Proceedings of the 10th Australian Space Science Conference Brisbane 27 - 30 September, 2010

Australian Space Science Conference Series

1st Edition
Published in Australia in 2011 by
National Space Society of Australia Ltd
GPO Box 7048
Sydney NSW 2001
Fax: 61 (02) 9988-0262
email: nssa@nssa.com.au
website: http://www.nssa.com.au

ISBN 13: 978-0-9775740-4-9

Editors: Wayne Short and Iver Cairns

Distributed on DVD

Preface to Proceedings

A large number of the presenters at the conference later submitted completed written papers which form the basis of the academic conference proceedings.

All papers published in these proceedings have been subject to a peer review process whereby a scholarly judgement by suitable individuals endorsed by the Program Committee determined if the paper was suitable to be published. In most cases papers were not initially accepted but were revised until deemed suitable.

The Editors would like to give special thanks to the Program Committee and those scholars who participated in the peer review process:

Brad Alexander, Jeremy Bailey, Mark Blair, Mark Bishop, David Buttsworth, Graham Brooker, Laurie Brown, Ben Cazzolato, Jonathan Clarke, Con Doolan, Alan Forghani, David Froning, Yue Gao, Jason Held, Charles Jenkins, Michael Jokic, Joseph Leach, Diai Liu, Kenneth Lynn, David Mee, David Neely, Marc Norman, Sean O'Byrne, Clare Paton-Walsh, Ian Petersen, Gordon Pike, Duncan Steel, Salah Sukkariah, Catherine Ticehurst, Matthew Tetlow, William Uther, Vincent Wheatley, David Willson, Lihua Zhao.

Finally we would like to thank our sponsors (the Space Policy Unit of DIISR and Engineers Australia) for their support to fund student participation and the Organising Committee for giving generously of their time and efforts. We trust that you will find the 2010 Conference Proceedings enjoyable and informative.

Wayne Short and Iver Cairns

Editors, 10ASSC Conference Proceedings

May 2011

Conference Background

The Australian Space Science Conference (ASSC) is the focus of scientific cooperation and discussion in Australia on research relating to space. It is a peer reviewed forum for space scientists, engineers, educators, and workers in Industry and Government.

The conference is of relevance to a very broad cross section of the space community, and therefore generates an enlightening and timely exchange of ideas and perspectives. The 2010 conference was set against the backdrop of the Australian Academy of Science and its National Committee for Space Science (NCSS) formally releasing the first Decadal Plan for Australian Space Science. The scope of the conference covers fundamental and applied research that that can be done from space and space-based platforms, and includes the following:

- **Space science**, including space and atmospheric physics, remote sensing from space, planetary sciences, astrobiology and life sciences, and space-based astronomy and astrophysics
- **Space engineering**, including communications, navigation, space operations, propulsion and spacecraft design, testing, and implmentation
- **Space industry**
- **Government, International relations and law**
- **Education and outreach**

The 10th ASSC was held at St Leo's College in the University of Queensland from September 27 to 30, 2010 . It consisted of a series of presentations by various researchers on topics that emphasize the diversity of endeavours and disciplines that encompass Space Science research in Australia. Appendix A has a copy of all abstracts submitted for presentation at the conference.

The 10th ASSC was organised by the National Space Society of Australia (NSSA) and the Academy of Sciences National Committee for Space Science (NCSS). The Australian Space Research Institute (ASRI) helped significantly.

A call for papers was issued in March 2010 and researchers were invited to submit abstracts for presentation at the conference. Following the conference itself, a call for written papers was issued in October 2010: this invited presenters to submit a formal written paper for this Proceedings that covered their abstracts.

Table of Contents

Table of Contents

Welcome to the 10th Australian Space Science Conference

and to St Leo's College, University of Queensland! This will be the fourth ASSC jointly sponsored and organised by the National Committee for Space Science (NCSS) and the National Space Society of Australia (NSSA). The ASSC is intended to be the primary annual meeting for Australian research relating to space science. It welcomes space scientists, engineers, educators, and workers in Industry and Government.

The launch of the first Decadal Plan for Australian Space Science on 27 September is a historic and crucial event for this ASSC, and for the development of Australia's space science community. Multiple workshops on implementing the Plan will take place.

This year are working for the first time with the Australian Space Research Institute (ASRI). We will also hear more about the successful projects from the first two rounds of the "Australian Space Research Program" (ASRP) – an initiative managed by the newly formed Space Policy Unit.

We look forward to an excellent meeting!

Iver Cairns
Co Chair ASSC 2010
Chair, NCSS

Wayne Short
Co Chair ASSC 2010
President, NSSA

The National Space Society of Australia is the coming together of like-minded space enthusiasts who share a vision for the future in which there is an ambitious and vigorous space program leading to eventual space settlement.

To this end the National Space Society (worldwide) promotes interest in space exploration, research, development and habitation through events such as science and business conferences, speaking to the press, public outreach events, speaking engagements with community groups and schools, and other pro-active events. We do this to stimulate advancement and development of space and related applications and technologies and by bringing together people from government, industry and all walks of life for the free exchange of information.

As a non-profit organisation, the National Space Society of Australia draws its strength from an enthusiastic membership who contributes their time and effort to assist the Society in pursuit of its goals.

For more information, and to become a member:

http://www.nssa.com.au

Ad Astra!
Wayne Short
NSSA President

The National Committee for Space Science (NCSS) is chartered by the Australian Academy of Science to foster space science, to link Australian space scientists together and to their international colleagues, and to advise the Academy's Council on policy for science in general and space science in particular. The associated web page can be reached at http://www.science.org.au/natcoms/index.htm . Accessible resources include the 2004-2006 Report on Australian Space Research.

NCSS believes that ASSC meetings provide a natural venue to link Australian space scientists and foster the associated science, two of its core goals. As well as ASSC, it is also sponsoring the VSSEC – NASA Australian Space Prize.

NCSS is the primary driver of the Decadal Plan for Australian Space Science. The plan is completed and was launched at this conference on Monday 27 September 2010. NCSS encourages all those interested in space science to help develop and support the Plan. Wishing you an excellent conference,

Russell Boyce (UQ), Iver Cairns (U. Sydney, Chair), Graziella Caprarelli (UTS), Alex Held (CSIRO, COSSA), Fred Menk (U. Newcastle), David Neudegg (IPS Radio Services), Bob Vincent (U. South Australia), Malcolm Walter (UNSW).

Decadal Plan: Launch & Workshops on Monday September 27

The launch of the first Decadal Plan for Australian Space Science, almost 5 years in development, was and remains a major priority for Australia's space scientists and for ASSC. Additional priorities are to bring together and educate the scientific community and associated stakeholders in Government and industry, and to hold workshops that start implementing the Plan.

The first morning session addressed the interests in Space of existing Government units. Presentations were be made by Space Policy Unit, IPS Radio & Space Services, the Bureau of Meteorology, and the Space Innovation Industry Council, among others.

The Decadal Plan was launched in the second morning session. The Chair of the NCSS opened the Launch with a brief summary of the context for the Decadal Plan. Then, Professor Peter Hall, the Academy's Secretary for Physical Sciences, described the importance of the Decadal Plan for Australia's space sector and welcomed Professor Margaret Sheil, CEO of the Australian Research Council to launch the Plan. Professor Sheil described the importance and successes of Australian space science, plus links to other areas, and launched the Plan. Subsequently the NCSS Chair provided a summary of the Decadal Plan and answered questions. The Launch was filmed by ABC TV and led to a 1-minute segment on the ABC's 6 o-clock News and many radio interviews by the NCSS Chair in Brisbane and Professor Malcolm Walter in Sydney.

The afternoon sessions started with the talk "Pathways to Space" on the successful ASRP grant, followed by Workshops on the Decadal Plan's Flagship projects "Spaceship Australis" and Marabibi "Constellation" and associated linkages, both international and domestic.

Additional Plan Workshops were held on 28 to 30 September, as in the program overview on page xv.

2010 ASSC Program Committee

Bill Barrett (Asia Pacific Aerospace Consultants)

Mark Blair (ASRI)

Iver Cairns (Chair, University of Sydney)

Jon Clarke (Mars Society of Australia)

Roger Franzen (Earthspace)

Brian Fraser (University of Newcastle)

Trevor Harris (DSTO)

Trevor Ireland (Australian National University)

Roman Makarevich (La Trobe University)

David Neudegg (IPS Radio Services)

Carol Oliver (University of New South Wales)

Michael Smart (University of Queensland)

Paul Tregoning (Australian National University)

Bruce Warrington (National Measurement Institute)

Colin Waters (University of Newcastle)

Xiaofeng Wu (University of Sydney)

2010 Organising Committee

Mark Blair
ASRI

Iver Cairns
Co Chair ASSC 2010
Chair, ASSC Program Committee

Sophia Casanova
Secretariat

Stuart Kearney
Australian Youth Aerospace Forum

Eloise Matheson
Secretariat

Wayne Short
Co Chair ASSC 2010
Project Manager, Operations

Michael Smart
University of Queensland

10th ASSC Conference Program
September 27 – 30

Time	Monday	Tuesday	Wednesday	Thursday
7:30				
8:00	Registration	Registration	Registration	Registration
		Decadal Plan Workshop *Leonian*	Decadal Plan Workshop *Leonian*	Decadal Plan Workshop *Leonian*
9:00	Government Space *Boardroom*	Plenary *Boardroom*	Plenary *Boardroom*	Space Technology & Engineering II *Boardroom* / Space Physics & Weather *Leonian*
11:00	Break	Break	Break	Break
11:30	Decadal Plan Launch *Boardroom*	Plenary *Boardroom*	Plenary *Boardroom*	Space Technology & Engineering III *Boardroom* / Astrobiology II *Leonian*
13:00	Lunch	Lunch	Lunch	Lunch
14:00	Decadal Plan Workshop *Boardroom* / *ASRI / Student Stream Leonian*	ASRP Session 1 Boardroom / Planetary & Astrobiology I Leonian	ASRP Session 2 Boardroom / Space Physics & Education Leonian	Closing Remarks & Feedback / Closed Workshop *Boardroom*
14:40				
15:30	Break	Break	Break	Close
16:00	Decadal Plan Workshop *Boardroom* / ASRI / Student Stream *Leonian*	Posters	Space Technology & Engineering I *Boardroom* / Earth Observation from Space & Planetary *Leonian*	
16:30				
18:00	Cocktail Reception	Close	Close	
18:15				
19:15	Close			

TWO NEW AUSTRALIAN SPACE DEBRIS OBSERVATORIES

Alan H Brockman[1], Myrtille Laas-Bourez[2] and John A Kennewell[3]

[1] *Learmonth Solar Observatory, IPS Radio and Space Services, PO Box 200, Exmouth, Western Australia, 6707*
[2] *University of Western Australia / International Centre for Radio Astronomy Research, M468, 35 Stirling Highway, Crawley, Western Australia, 6009*
[3] *Australian Space Academy, PO Box 3, Meckering, Western Australia, 6405*

Summary: Orbital space debris is a current and continuing issue with regard to space operations, as evidenced by the adoption of mitigation guidelines by the United Nations General Assembly in 2007. One of the primary tasks of mitigation is the tracking and study of the populations of space objects in both low Earth and geosynchronous orbit. Although Australia has not produced a large number of space objects, either active satellites or debris, all nations will eventually be affected by space debris. Australia resides at a longitude and latitude that has been poorly served by ground-based space surveillance assets. Two independent facilities have recently been set up to help rectify this deficiency, one at Learmonth, and one at Gingin, both in Western Australia. We describe these systems and their place in contributing to our knowledge of the overall population of orbital space debris.

Keywords: orbital space debris, Australia, HANDS, Zadko, optical monitoring

LSO – HANDS

For many years the Ground-based Electro-Optical Deep Space Surveillance (GEODSS) system has been the backbone of the US Space Surveillance Network optical sensors. GEODSS is currently deployed at three sites around the world. To augment this system, a network of smaller telescopes is now being deployed worldwide, under the acronym High Accuracy Network Determination System (HANDS). One of these systems has recently been established at the Learmonth Solar Observatory (LSO) on the North West Cape of Western Australia (22. ^{o}222 S, 114. ^{o} 101 E).

This system was established under an agreement between the Australian Bureau of Meteorology and the US Department of Defense. It consists of two automated 40 cm telescopes with associated weather stations, data processing and data communications equipment. All operations are tasked from a US Air Force Research Laboratory AMOS (Air Force Maui Optical Station) control facility on the island of Maui in Hawaii. The processed data is also returned to this centre.

The equipment at LSO is housed in two astronomical domes from Ash Manufacturing Company and one data processing and communications shed. All three buildings are air conditioned, as the summer temperatures at Learmonth can reach nearly 50°C. The dome shutters, however, open to the atmosphere at night if the weather stations (of which there are two for redundancy) indicate suitable conditions. The system operates entirely autonomously under tasking from the AMOS centre.

Fig 1 Learmonth Solar Observatory showing the location of the HANDS domes.

Fig 2 Close-up of the two HANDS domes. North is to the right of the image.

The northern dome houses a 40 cm f/9 Ritchey-Chrétien reflecting telescope from RC Optical Systems in Flagstaff, Arizona. This is mounted on a Paramount equatorial mount from Software Bisque in Golden, Colorado. This telescope is used for high-accuracy observations of space objects (active satellites and space debris) in geosynchronous and near-geosynchronous orbital space, and is part of the HANDS proper. The CCD camera at the focal plane of this telescope is an Apogee U47 with 1024x1024 13mm pixels giving a 0.31 degree FOV and a plate scale of 1.09 arc-second per pixel.

The telescope in the south dome is a 40 cm MOSAIC (Maui Optical System and Imaging Center) Mstar modified f/4.76 Cassegrain reflector coupled to a 2048x2048 13.5mm pixel Fairchild CCD camera giving a 1.2 degree field of view and a plate scale of 1.49 arc-second per pixel. This telescope is also mounted equatorially on a Paramount, and is used primarily for fast moving space objects.

Fig 3 The northern telescope used for GEO objects.

Fig 4 The southern telescope used for LEO and other objects.

Zadko telescope and TAROT network

The Zadko telescope was made possible by a bequest from Jim Zadko and is owned and operated by the University of Western Australia. It is co-located with the Australian International Gravitational Observatory at Gingin, 70 km north of Perth (31.^{o}357 S, 115.^{o}714 E).

The Zadko telescope is a 1m f/4 Cassegrain reflector constructed by DFM Engineering Ltd. The fork mount has a slew rate of 3 degrees per second. The telescope feeds an Andor DW 436 camera with a Marconi 42-40 back illuminated 2048 x 2048 CCD sensor. The field of view is 23.6' x 23.6' and the plate scale is 0.69 arcseconds per pixel. This corresponds to a linear distance of only 150 metres at the height of geosynchronous orbit. A 180 second exposure reaches a limit of 21st magnitude at 3 sigma above the noise.

Fig 5 (left) An aerial view of the Gingin observatory indicating the location of the Zadko telescope dome. Fig 6 (above) Ground view of the Zadko telescope.

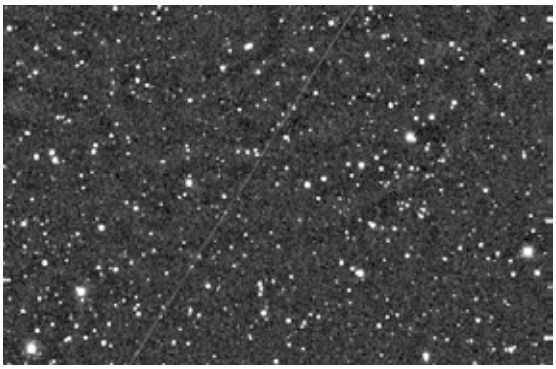


Fig 7 The image above shows the trail of the now defunct Australian Fedsat satellite as it moves in low Earth orbit.

Fig 8 The location of the Zadko telescope and the other two telescopes in the TAROT network.

The Zadko telescope became fully functional in April 2010, and has now been incorporated into the French TAROT (Télescope à Action Rapide pour les Objets Transitoires / Rapid Action Telescopes for Optical Transients) network, which also includes telescopes in France and Chile. These three telescopes provide good longitudinal coverage of the geosynchronous belt. Approximately 15% of observing time of the network is allocated to space object identification and tracking. Automated algorithms allow 1000 measurements and reductions per night.

Summary

The HANDS system at Learmonth provides metric and photometric data on satellites and space debris from low Earth to geosynchronous orbit. The data is incorporated into the US Space Object Catalog which is available to registered users at *www.space-track.org*.

The Zadko telescope will provide data that will contribute to the growing space situational awareness of the European Space Agency through the French Space Agency CNES (Centre National d'Études Spatiales). The data could be made available to an Australian organisation for conjunction assessment of Australian space assets. It is believed that the one metre aperture of the telescope will allow detection of space objects to below 0.1 metre in size at geosynchronous distances.

References

[1] Coward, D, et al, "The Zadko Telescope: A Southern Hemisphere Telescope for Optical Transient Searches, Multi-Messenger Astronomy and Education", *Publications of the Astronomical Society of Australia*, **27**, 331-339, 2010.

[2] Laas-Bourez, M, G Blanchet, M Boer, E Ducrotte & A Klotz, "A New Algorithm for Optical Observations of Space Debris with the TAROT Telescopes", *Advances in Space Research*, **44**, 1270-1279 (2009).

[3] Sabol, C, S Carter & D Vallado, "A Fresh Look at Angles-Only Orbit Determination", *AAS/AIAA Astrodynamics Specialist Conference*, AAS-93-363, 1999.

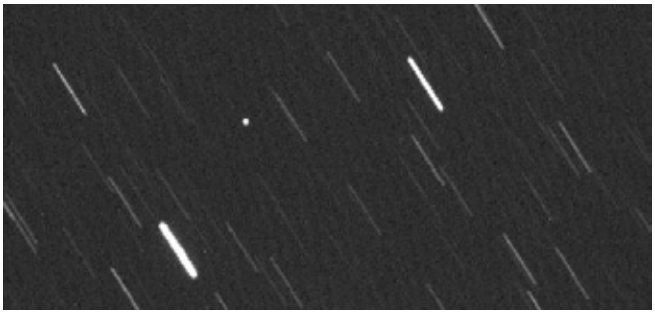


Fig 9 The first light image from the HANDS system of a geosynchronous satellite (left of centre). The telescope is pointed in a fixed direction which causes the stars to show as trails, as the Earth rotates beneath them.

Some Observations and Analysis of Australian Space Debris

Myrtille Laas-Bourez[1], John A. Kennewell[2] and David M Coward[3]

1 University of Western Australia / International Centre for Radio Astronomy Research M468, 35 Stirling Highway, Crawley, W.A., 6009, Australia
2 Australian Space Academy, PO Box 3, Meckering, W.A., 6405, Australia
3 University of Western Australia / School of Physics M013, 35 Stirling Highway, Crawley, W.A., 6009, Australian Space Academy

Summary: We describe observations of Australian space debris made with the Western Australian one metre Zadko Telescope. Analysis of positional data from these observations is compared with standard US Strategic Command orbital elements. Discrepancies are discussed, and a brief analysis of the long-term evolution of an Australian geosynchronous debris object is presented. Finally, we discuss Australian space debris orbits with reference to current international space debris mitigation guidelines.

Keywords: Zadko Telescope, space debris, image processing, Australian GEOSATS.

Introduction

A total of 14 Australian satellites have been launched into orbit over a 43-year period. Ten of them are geosynchronous communication satellites. One of them failed to deploy from its upper stage and thus never made it to geosynchronous orbit; it decayed into the Earth's atmosphere together with its booster rocket after 2.5 years. Of the remaining nine, four reached the end of their useable life and were boosted into supersynchronous orbit. To remain clear of active GEOSATS, these graveyard orbits need to be at least 250 km, and preferably 300 km, above true geostationary orbit. This allows for maneuvering of active satellites and for natural orbit perturbations of defunct satellites. Although the risk of collision is small, monitoring space objects existing in, or passing through, geosynchronous space is desirable. Ground based observations are currently limited to the larger objects because of the distance involved and larger aperture telescopes, such as the one described here, have the potential to explore smaller size populations.

The Western Australian one metre Zadko Telescope (ZT) has recently been involved in a program that tracks geosynchronous space debris. This fully robotic optical facility is the Australian node of a global robotic telescope network. Two of the three telescopes of this network are employed to track geosynchronous satellites by the French national space agency CNES (*Centre Nationale d'Études Spatiale)*, and are involved in a survey campaign for the European Space Agency (ESA). The ZT will soon participate in these observations. In this paper, we first discuss the international robotic telescope network followed by some observations of Australian space objects acquired using this network. We compare our measurements with readily available orbital data in Two Line Elements (TLE) format from the United States Strategic Command. We conclude with a discussion of Australian space debris.

Telescope Network Description

The TAROT *(Télescope à Action Rapide pour les Objets Transistoire* – Rapid Action Telescope for Transient Objects*)* telescopes (http://tarot.obs-hp.fr/) are fully robotic 25 cm telescopes. One is located in France and the second in Chile. Both telescopes have for their primary goal the observation of the prompt optical counterparts of Gamma-Ray Bursts (GRBs) [1,2]. The two TAROT telescopes are optically and mechanically identical (Table 1). More details are given by Klotz et al [3].

The ZT is a one-metre f/4 Cassegrain telescope constructed by DFM Engineering Ltd and is situated in Gingin, 70 km north of Perth, Western Australia (Table 1) [4]. It is an equatorial fork-mounted Cassegrain telescope with a primary mirror aperture of one metre and a focal length of four metres.

The robotisation of each telescope requires software modules to manage security, observation planning, image processing, data archiving and coordination of these functions. The last major update of the software occurred in January 2009. A detailed description of the software can be found in Klotz et al [3]. For the two TAROT telescopes and the Zadko telescope, all these processes are fully automated, from scheduling of observation requests to the processing of data.

Table 1: *Technical description of the TAROT telescopes and the ZT.*

Telescope	TAROT Calern	TAROT Chile	ZT
GPS coordinates	43.75°N 6.92E 1320m	29.26°S 70.73°W 2398m	31.36°S 115.71E 50m
Primary mirror	25 cm		1 m
Mount	Equatorial fork mount		
Focal length	0.85 m		4.04 m
Field of view	1.86° x 1.86°		23.6’ x 23.6’
Max slew speed	60° /sec		3° / sec
Camera CCD	camera Andor DW436 / Marconi CCD 42-40 back illuminated		
CR limit at 3-σ	18.2 with 30 sec		21 with 180sec
Fully robotised in	1999	2006	2010

Table 1 provides technical descriptions of the three telescopes. More details can be found in references [2,3,4].

The real advantage of this robotic telescope network is the geographic distribution of the telescopes. Together they provide the opportunity to scan and track objects across a large part of the sky. Nearly all geosynchronous belt objects can be observed by the network. Only those in a 20-degree longitude span above the middle of the Pacific Ocean are not visible (shaded area in Figure 1). This is due to both telescope horizon limitations and site positioning.

Approximately 15% of TAROT observing time is dedicated to space debris identification and tracking led by CNES. Accurate, rapid and robust methods have been developed for near real-time detections and enable systematic surveys of both known and unknown satellites and debris [5]. A GEO observation strategy with the ZT is similar to that used by the TAROT telescopes. Three images, each of a few

seconds duration, are taken of the same star field without sidereal tracking. Geostationary objects appear as point sources, while stars are line segments due to trailing caused by the Earth's rotation. Three images are required for the algorithm that automatically reduces the object positional data from the image as well as for discriminating against cosmic ray hits (which appear on only one image). This latter source of interference becomes more important at higher site altitudes, and less important for the ZT sea level site.

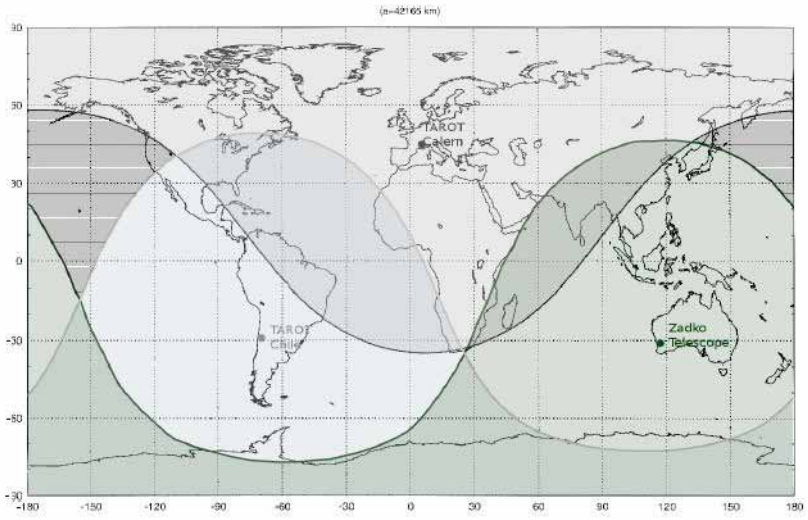


Figure 1: *The GEO belt coverage of the three-telescope network. The black and white hatched area over the northern Pacific is the only GEO region not seen by the network.*

Australian space objects

Australia is a relatively minor contributor to satellite launches, the creation of space debris, and in space situational awareness, including the surveillance of artificial space objects. In general, space object tracking facilities in Australia have involved the hosting of equipment from other countries. This includes a new space track facility at Learmonth Solar Observatory by the US Air Force Research Laboratory as part of the US Space Surveillance Network. A review of some other facilities and possibilities for future Australian activities in this area has been made by Newsam [6].

In Australia, Electro Optic Systems (EOS) is building an active space tracking system using automated laser ranging equipment to track small space objects in low Earth orbit (LEO). The Australian Space Research Program is supporting EOS Space Systems to automate an existing manned tracking system to reduce operating costs. This project aims to provide high-precision laser and optical tracking to improve space situational awareness. The system is expected to be able to detect objects to below ten centimetres in size at distances up to 1000 km [7]. As this is a lidar system with a sensitivity dependence on the inverse fourth power of the target distance, it is not suitable for GEO observations.

The 14 Australian satellites that have been launched into orbit are described in Table 2. Ten of them are geosynchronous communication satellites, the other four have been launched into low Earth Orbit. As well as these satellites, at least four rocket

bodies used to launch the GEOSATs are still in orbit. These are Optus A1 R/B (PAM-D), Optus A2 R/B(PAM-D), Optus B1 R/B(Star 63F) and Optus B3 PKM (see the last four entries in Table2). All four objects are in Geostationary Transfer Orbits (GTO), and cross orbital regimes from LEO space to GEO space, therefore posing a potential hazard to all space objects, both active and debris, in these orbits.

Table 3 shows apogee, perigee and orbital inclination for non-operational Australian satellites and rocket bodies as indicated by the USSTRATCOM satellite catalog on the 9th of September 2010. Four rocket bodies are still in GTO and four satellites are now classed as debris in supersynchronous orbits. Two objects add to the LEO space debris field.

Table 2: *Listing and status of the 14 Australian satellites that have been launched into orbit to date. The last four entries are for rocket bodies still in orbit.*

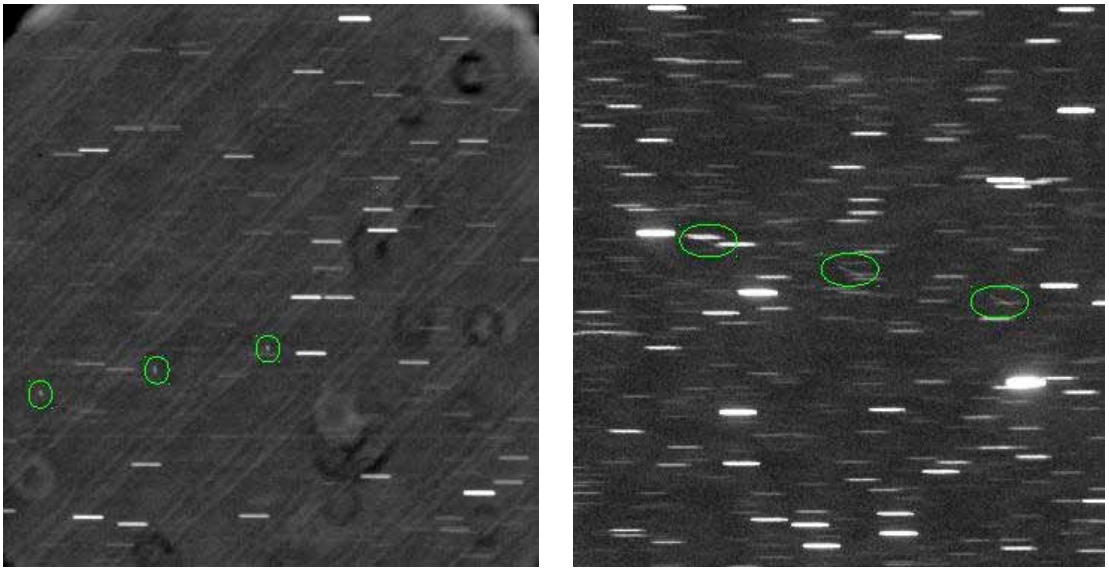
Satellite Name	**International Designation**	**Launched**	**Orbit**	**Purpose**	**Status 2010**
WRESAT	1967-118A	1967/11/29	LEO	Scientific	Decayed 68/01/10
OSCAR 5 (AO-5)	1970-008B	1970/01/23	LEO	Amateur / Sci	Debris in orbit
Optus A1 (AUSSAT 1)	1985-076B	1985/08/27	GEO	Commsat	Debris above GEO
Optus A2 (AUSSAT 2)	1985-109C	1985/11/27	GEO	Commsat	Debris above GEO
Optus A3 (AUSSAT 3)	1987-078A	1987/09/15	GEO	Commsat	Debris above GEO
Optus B1 (AUSSAT B1)	1992-054A	1992/08/13	GEO	Commsat	Debris above GEO
Optus B2 (AUSSAT B2)	1992-090A	1992/12/21	GTO	Deploy fail	Decayed 95/06/29
Optus B3	1994-055A	1994/08/27	GEO	Commsat	Operational
WESTPAC	1998-043E	1998/07/10	LEO	Laser reflector	Operational
FEDSAT	2002-056B	2002/12/04	LEO	Scientific	Debris in orbit
Optus C1	2003-o28B	2003/06/11	GEO	Commsat	Operational
Optus D1	2006-043B	2006/10/13	GEO	Commsat	Operational
Optus D2	2007-044A	2007/10/05	GEO	Commsat	Operational
Optus D3	2009-044B	2009/08/21	GEO	Commsat	Operational
Optus A1 R/B (PAM-D)	1985-076E	1985/08/27	GTO	Launcher	Debris in GTO
Optus A2 R/B (PAM-D)	1985-109G	1985/11/27	GTO	Launcher	Debris in GTO
Optus B1 R/B (Star 63F)	1992-054C	1992/08/13	GTO	Launcher	Debris in GTO
Optus B3 PKM	1994-055C	1994/08/27	GTO	Launcher	Debris in GTO

Table 3: *Apogee and perigee heights and orbital inclination for non-operational Australian satellites or rocket bodies. (Source: USSTRATCOM SATCAT 9 September 2010).*

Space Object Name	**Perigee Ht (km)**	**Apogee Ht (km)**	**Inclination (deg)**
Optus A1 (Aussat 1)	35900	35962	12.92
Optus A1 R/B (PAM-D)	333	34829	25.73
Optus A2 (Aussat 2)	35869	35894	12.69
Optus A2 R/B (PAM-D)	430	34829	25.9
Optus A3 (Aussat 3)	36149	36197	11.35
Optus B1 (Aussat B1)	36042	36132	3.91
Optus B1 R/B (Star 63F)	350	36690	22.7
Optus B3 PKM	412	38315	23.3

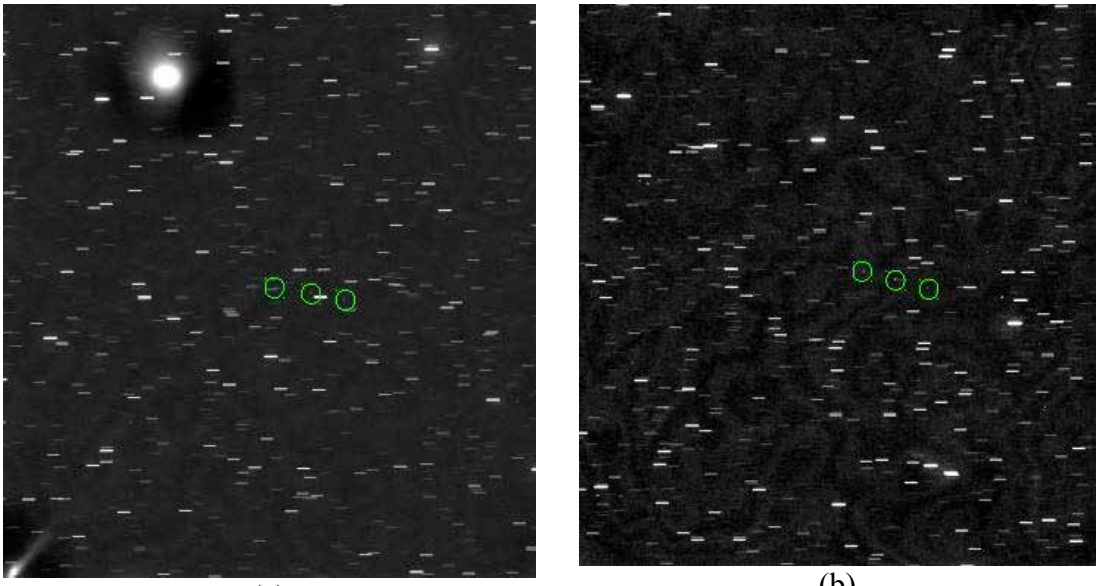
Images of several Australian space debris were made with the ZT and the TAROT network, and these were automatically reduced by the robotic software to produce

positional coordinates. The images shown in figures 2 and 3 are the sum of three exposures for each of Optus A1, Optus A3 and the Optus B3 rocket body. All these bodies are inclined with respect to the equatorial plane, with inclinations ranging from 11 to 23 degrees.



(a) (b)

Figure 2: *(a) Sum of three ZT images of Optus A3 (highlighted by green circles) made around 10:25 UTC on 2010/08/04. (b) Sum of three part images of Optus B3 PKM (green ovals) made around 00:00 UTC on 2010/08/10.*



(a) (b)

Figure 3: *(a) Sum of three TAROT Chile images of Optus A1 (green circles) made around 03:30 UTC on 2010/08/09. (b) Sum of three TAROT Chile images of Optus A1 (green circles) made around 07:30 UTC on 2010/08/09.*

Positional Data

Current observations using the ZT have a time accuracy of around one second. The TAROT instruments have time stamping accuracy in the millisecond range, which

includes a correction for CCD shutter opening variations across the aperture. For GEO objects observed without sidereal tracking, a Gaussian is the most commonly used model for the image shape in order to extract the object position from the image. However, in supersynchronous or GTO orbits objects will be spread or trailed on the image. Using a Gaussian fitting, we can extract the centre of the trailed image with a one-pixel accuracy. In this configuration, measurements made with the ZT have an uncertainty in the tangential direction of:

- 3.2 km in the GEO belt
- ~0.5 km for LEO objects (200 to 1500 km orbital heights)

Millisecond time stamping for the ZT is currently being implemented and when completed satellite positional error to 0.5 pixel should be on the order of:

- 0.07 km in the GEO belt
- ~0.01 km for LEO objects

The new time stamping hardware will be accompanied with a new fitting method, adapted for supersynchronous and GTO objects, with regard to the image shape (using a method adapted from the one presented by Lass-Bourez et al [8]).

In September 2010, observations of some Australian GEOSATS were made with the ZT and TAROT network. Each image was astronomically calibrated with the Tycho-2 star catalogue [9] and the TAROT image-processing pipeline. All positions were extracted with a Gaussian image fit. Table 4 lists some of these measurements for Optus A1 and Optus A3.

Table 4: *Comparison between the astronomical coordinates (RA and DEC J2000 in degrees) extracted from the observed satellites and the ones computed with the nearest TLE from the USSTRATCOM catalogue. RA_{TLE} and DEC_{TLE} are the coordinates computed from the TLE. The altitude h_{TLE} of the object is that computed from the TLE (mean motion conversion to semi-major axis) using a terrestrial equatorial radius of 6378 km. RA and DEC are the coordinates extracted from the images. ΔRA and ΔDEC are the differences between the two different estimates.*

Object	**Telescope**	**Date UTC**	**RA_{TLE} (deg)**	**DEC_{TLE} (deg)**	**h_{TLE} (km)**	**RA(deg)**	**DEC (deg)**	**ΔRA (km)**	**ΔDEC (km)**
Optus A3 TLE updated 2010/8/1	ZT	2010/8/4 10:24:56	263.092	-1.9457	36175	263.039	-1.9849	34	25
		2010/8/4 10:25:17	263.175	-1.9612		263.123	-2.0000	34	25
		2010/8/4 10:25:37	263.258	-1.9768		263.206	-2.0155	34	25
Optus A1 TLE updated 2010/8/5	TAROT Chile	2010/8/9 03:30:04	348.008	-6.7486	35863	347.949	-6.7640	38	10
		2010/8/9 03:30:33	348.127	-6.7307		348.067	-6.7468	38	10
		2010/8/9 03:31:01	348.244	-6.7129		348.184	-6.7295	39	11
Optus A1 TLE updated 2010/8/5	TAROT Chile	2010/8/9 07:29:57	46.771	6.2496		46.697	6.2348	50	10
		2010/8/9 07:30:25	46.886	6.2782		46.811	6.2633	51	10
		2010/8/9 07:30:53	46.999	6.3067		46.925	6.2906	50	11

In the above table, differences in angular coordinates are converted into linear distances (plane of sky) using a scale factor $E = 2 \pi r / 360$ (km/deg), where r is the distance between the observing site and the target space object.

The large discrepancy, up to 50 km in right ascension, between the two sets of data is rather disconcerting, and has been verified as existing in similar comparisons. The potential accuracy in the ZT and TAROT optical observations is at least an order of magnitude better than this. At geosynchronous altitudes the perturbing forces do not cause rapid changes in the orbital elements. The date/time difference between the observations that led to the TLE generation and the optical observations reported here is not expected to contribute significantly to the positional discrepancy. Further work needs to be done to trace the source of this disagreement. In general, passive optical observations are expected to provide the most accurate positional data because of the inverse square law dependence, versus the inverse fourth power law for active sensors.

Orbital stability of Australian GEO space debris

A study was undertaken to examine the stability of an Australian defunct communications satellite that is now in supersynchronous orbit. The Optus A1 satellite was chosen for orbit propagation over a 200 year interval, starting with the USSTRATCOM orbital elements dated 9 September 2010. The study was performed for us by the European Space Agency, and the results are indicated graphically in Figure 4. Four different methods of propagation were employed as a check on each other, all showing very similar orbital evolution. The first method, from Cowell [10], allows direct integration of second-order equations from orbital dynamics. A second method used the GBS (Gragg-Bulirsh-Stoer) algorithm where step and order are variables (black points in Figure 4 – calculation time 2 hours) [11]. The third method was a semi-analytic technique using the TheoNA (*Théorie Analytique Numérique* – Numeric Analytical Theory) software (red points in Figure 4 – calculation time 2 minutes) [12]. The fourth and last method was another semi-analytical technique using the CODIOR (*COrrection Différential d'ORbite* - Differential Correction of Orbits) CNES software [13].

The key conclusions of this study are that the semi-major axis (with an amplitude of oscillation of only 8 km) and the eccentricity (varying between 0.00001 and 0.00135) are quite stable. All four techniques agree on this stability. The orbital inclination does show an approximate 50-year periodicity with an amplitude of 15 degrees. Higher inclination orbits pose a slightly greater collisional hazard.

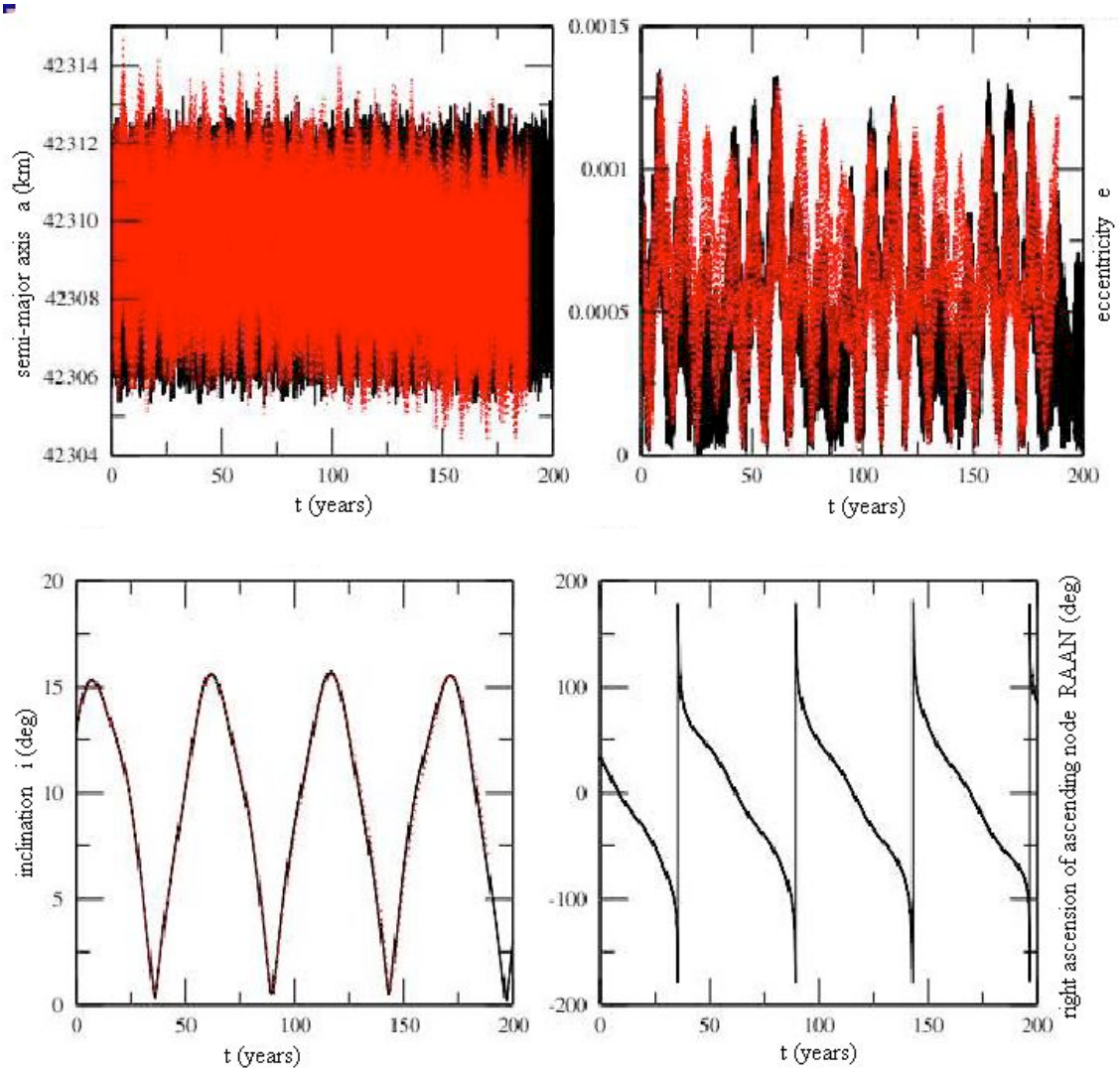


Figure 4: *Long term evolution of the defunct Australian satellite Optus A1. Implementation of several integration techniques: black traces show the numerical method of GBS (Gragg-Bulirsh-Soer) whereas red traces show the semi-analytical method using the TheoNA software.*

Summary and discussion

Unlike LEO objects that are subject to atmospheric drag (which only really becomes significant below altitudes of ~700 km), geospace objects experience no significant atmospheric forces. The perturbing forces in this region are gravitational (terrestrial non-sphericity, solar and lunar) and solar radiation.

To keep non-operational satellites clear of active geostationary satellites, they are normally boosted into supersynchronous orbits (i.e. orbits above geostationary orbit) at the end of their lifetime. It is now generally recognised that geospace disposal orbits should be at least 250 km, and preferably 300 km, above the geostationary orbit which has an altitude of 35787 km (above the Earth's equator). This means that the perigee of a disposal orbit should not lie below 36037 km. This specified height increment above geostationary orbit allows for the normal movement of active

geosynchronous satellites (due to perturbations) within a specified station-keeping box, or volume of space.

Five Australian space objects (debris) do not meet these current guidelines. Optus A1 and Optus A2 are between 100 and 200 km above the geostationary orbit. Optus B1 has its perigee around 250 km above the geostationary orbit. The rocket bodies Optus B1 R/B (Star 63F) and Optus B3 PKM are in transfer orbits that cross the geostationary orbit. It should be noted that Optus A1 and A2 were placed into their disposal orbits well before the issue of geospace debris was fully appreciated and before the United Nations General Assembly passed orbital debris mitigation guidelines, the seventh and last of which pertains to geospace debris.

A space debris survey consists of observation, measurement, orbit determination, catalogue update and forward orbit propagation. The US Space Surveillance Network currently tracks and provides orbital information for Australian space objects (both active and defunct). The risk of collisional fragmentation is currently very small in geospace, several orders of magnitude below that for low Earth orbiting objects, and dominated by the natural meteoroid environment that impacts the Earth. However, the European Space Agency will be using orbital positions provided by the Zadko Telescope to assist its growing space situational awareness program. This program will ultimately be engaged in independently cataloguing and performing conjunction assessments for all space objects from LEO to GEO. Positional discrepancies between the different data sources need to be investigated. Importantly, continued monitoring with higher positional precision can be used to establish baselines and monitor changes and fragmentations, whether from explosive or unlikely collisional events.

The motivation for the observations and analysis presented here is to bring the issue of Australian space debris to a wider audience.

Acknowledgements

M. Laas-Bourez thanks D. Hautesserres from CNES for the studies of Optus A1 presented in this paper (Figure 4).

D. M. Coward is supported by an Australian Research Council Future Fellowship The ZT, made possible by a philanthropic donation to UWA, is managed by the School of Physics, UWA, the International Centre for Radio Astronomy Research (ICRAR), and the Australian International Gravitational Research Centre (AIGRC).

References

[1] Boër, M., Atteia, J. L., Damerdji, Y., Gendre, B., Klotz, A. and Stratta, G., "Detection of a very bright optical flare from the Gamma-Ray Burst GRB 050904 at Redshift 6.29", *Astrophys. J.*, **638**, L71-L74, 2006.
[2] Klotz, A., Boër, M., Atteia, J. L. and Gendre, B., "Early optical observations of GRB sources by the TAROT telescope: period 2001-2008", *Astron. J.*, **137**, 4100-4108, 2009.

[3] Klotz, A., Boër, M., Eysseric, J., Damerdji, Y., Laas-Bourez, M., Pollas, C. and Vachier, F., "Status of the robotic TAROT observatories: Period 2004-2007", *Publications of the Astronomical Society of the Pacific*, **120**, 1298-1306, 2008.
[4] Coward, D. M., Todd, M., Vaalsta, T. P., Klotz, A., Zadko, J., Laas-Bourez, M., et al, "The Zadko Telescope: A southern hemisphere telescope for optical transient searches, multi-messenger astronomy and education", *Publications of the Astronomical Society of Australia*, **27**, 331-339, 2010.
[5] Laas-Bourez, M., Blanchet, G., Boër, M., Ducrotté, E. and Klotz, A., "A new algorithm for optical observations of space debris with the TAROT telescopes", *Adv. Space Res.*, **44**, 1270-1279, 2009.
[6] Newsam, G., "Surveillance of Space in Australia", *Proc. of the Advance Maui Optical and Space Surveillance Technologies Conference*, held in Wailea, Maui, Hawaii, September 17-19, 2008.
[7] Smith, C., "The EOS Space Debris Tracking System", *Proc. of the Advanced Maui Optical and Space Surveillance Technologies Conference*, held in Wailea, Maui, Hawaii, September 17-19, 2006.
[8] Laas-Bourez, M., Klotz, A., Blanchett, G., Boër, M. and Ducrotté, E., "Algorithms improvement in image processing for optical observations of artificial objects in geostationary orbit with the TAROT telescopes", *Proc. SPIE 7000*, **74**, 1-12, 2008.
[9] Hog, E., Fabricus, C., Makarov, V. V., Urban, S., et al., "The Tycho-2 catalogue of the 2.5 million brightest stars", *A & A*, **355**, L27-L30, 2000.
[10] Van Dooren, R., "Stabilization of Cowell's classic finite difference method for numerical integration", *Journal of Computational Physics*, **16**, 186-192, 1974.
[11] Deuflhard, P., "Order and stepsize control in extrapolation methods", *Numerische Mathematik*, **4**, 399-422, 1983.
[12] Golikov, A., "THEONA theory of Relative Satellite Motion Flying in Formation", *Proc. of the 18th International Symposium on Space Flight Dynamics (ESA SP-548)*, p.59, Munich, Germany, October 11-15, 2004.
[13] Exertier, P., Deleflie, F. and Métris, G., "Théorie du Mouvement Orbital Moyen: l'outil Informatique CODIOR", *Observatoire de la Côte d'Azur, Technical publication*, 2003.

Hypersonic inlet for a laser powered propulsion system and its interaction with an idealized laser induced detonation wave

Alan Harrland*, Dr. Con Doolan* and Dr. Vincent Wheatley†

* *School of Mechanical Engineering, The University of Adelaide, SA, Australia, 5005*
† *Centre for Hypersonics, The University of Queensland, QLD, Australia, 4072*

Summary: The idea of laser powered lightcraft was first conceptualised in the early 1970's as a means of launching small scale satellite payloads into orbit at a much lower cost in comparison to conventional techniques. Recently interest has been renewed in this innovative technology. Propulsion is produced via laser induced detonation of an incoming hypersonic air stream. This process requires suitable engine configurations that offer good performance over all flight speeds and angles of attack to ensure the required thrust is maintained. Stream traced hypersonic inlets have demonstrated the required performance in conventional hydrocarbon fuelled scramjet engines, and has been applied to the laser powered lightcraft vehicle. This paper will outline the current methodology, with a particular focus on the interaction between the laser induced detonation wave and the resulting inlet. An idealised computational fluid dynamics simulation has been produced to approximate the thrust produced by the detonation and it's interaction with the vehicle. The performance of the lightcraft inlet has also been evaluated.

Keywords: Hypersonic inlet design, lightcraft, laser powered propulsion, computational fluid dynamics

Introduction

The idea of Laser Propelled Lightcraft Vehicles was first conceptualised in the early 1970's[1] as a means of achieving low cost earth to orbit payload launches. Since 1986 the U.S Security Defence Initiative Organisation has supported research programs in Laser Propulsion systems. Laser Propelled Lightcraft Vehicles have also been investigated by the Air Force Research Laboratory (AFRL)[2]. Power is transmitted to the Lightcraft from a ground based laser, and air is utilised during flight in the sensible atmosphere. When the Lightcraft exits this region of usable atmosphere, the laser is then used to heat an ablative fuel source providing anaerobic propulsion.

Current conceptual and experimental designs of Lightcraft are explained in [2]. This design consists of three main sections; the forebody, the engine cowl and the afterbody, as indicated in Figure 1. The conical forebody acts as both an aerodynamic shape for providing lift to the craft, and a supersonic ramp to compress the incoming air before it enters the engine cowl for laser induced detonation. The engine cowl acts as both an inlet and an impulsive thrust surface for the detonation wave. The parabolic afterbody acts as both a primary receptive optic for the incident laser beam, and an expansion nozzle for the heated exhaust flow.

Thrust is achieved by a repetitively pulsed laser beam fired from a ground based laser. Plasma ignition is created by the laser-induced breakdown of gas in the engine cowl [3]. This

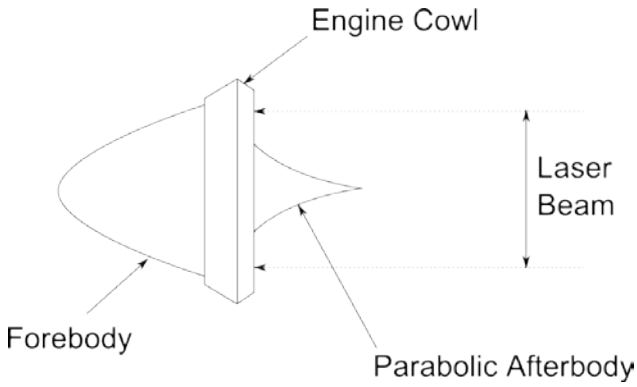


Fig. 1: Current lightcraft configuration

breakdown creates what has been labelled a detonantion wave that exerts a propulsive force on the engine cowl, generating the required thrust for the vehicle. Although there is no actual combustion 'per se', the terminology 'detonation wave' is employed in this paper to keep consistent with other literature.

The current work involves the design of a supersonic inlet for the lightcraft. It is essential to ensure the laser supported detonation process is provided with working gas at the required conditions throughout the flight envelope. If the inlet flow is un-started by the propulsion system, thrust requirements of the mission will not be met.

Laser supported detonation wave

Thrust is generated in the lightcraft vehicle by focusing a high powered laser beam adjacent to a solid surface. When the laser beam is intensified, the working gas is broken down and a detonation wave is formed. This wave then expands, imparting momentum on the adjacent surface, thereby generating thrust[3]. Either a steady state or continuously working laser pulse can be employed in thrust generation, however it has been found that the plasma created by a continuously working laser propulsion system is inherently unstable [4]. By pulsing the laser at a duration that is sufficiently short, the thrust generation can be highly efficient and much more stable. Due to the disruptive nature of the ionisation and plasma generation process, the laser induced breakdown has been likened to a nuclear explosion, or fireball[5]. It is highly important to develop an understanding of the effect that the laser detonation has on the inlets performance, as the two processes are intrinsically linked.

The pulsed laser induced detonation process is a highly complex process, involving many different physical phenomena occurring during the intermittent laser pulses. Quantum electro-dynamics, optics, fluid mechanics, gas dynamics and high temperature plasma dynamics all play a role in the laser induced detonation wave formation [6]. The laser induced detonation process can be divided into four successive stages[5], all of which occur in the time between pulses:

1) The initial air breakdown, where ionisation occurs in the cold gas and the initial plasma appears
2) Interaction between the remainder of the laser pulse energy and the initial plasma gas

3) Formation of the detonation process, where the blast wave relaxes in an unpowered manner across the thrust surface.
4) Extinction of the detonation wave

When the focused energy of the pulsed laser beam is higher than the threshold value of the gas, an initial optical breakdown stage occurs. The time taken for the initial breakdown and formative growth stages to occur is typically very small, the duration often less than a few nanoseconds. The thrust generated by this stage is small, and can be considered insignificant in the total thrust produced by the craft. The molecular working gas is broken down into an ionised plasma via two processes; cascade ionisation (the Inverse Bremsstrahlung effect) and multi-photon ionisation. In cascade ionisation, free electrons absorb electromagnetic radiation thereby increasing their kinetic energy. The electrons undergo random oscillating motion, causing collisions with atoms of the working gas. If the electron has sufficient kinetic energy, neutral particles can be ionised by collision with the free electrons, as shown by Eqn. 1[6].

$$e^- + M \rightarrow 2e^- + M^+ \tag{1}$$

The resulting free electrons have a reduced kinetic energy, which is then again increased by the laser energy. The kinetic energy of the free electrons are once again raised to a point sufficient to ionise another atom, causing a 'cascade' effect of the number of free electrons available. If the gas pressure, flash duration and laser intensity are sufficiently large, gas breakdown is caused by this ionisation[5]. If the intensity of the beam is not strong enough, loss processes slow and prevent the cascade effect.

The second process, multi-photon ionisation, increases the number of free electrons through the ionisation of particles by the absorption of photons. When particles absorb photon energy above the particle ionisation energy, electrons are caused to detach from the atom. This process is described in Eqn. 2. The multi-photon ionisation process requires significantly higher laser intensities than cascade ionisation, and as such plays a much smaller role in the generation of the plasma cloud.

$$M + mh\nu \rightarrow e^- + M^+ \tag{2}$$

The two ionisation processes occur at a faster rate than the pulse rate of the laser, and finish while the laser pulse is still focused on the plasma. This results in additional energy being added to the highly ionised, conducting hot expanding plasma[3]. Free electrons in the working gas are continually accelerated and the plasma begins to move up the laser beam until the pulse ends. During this period multiple plasmas may interact and combine. A blast wave front exists due to the high amount of energy absorbed. The front then expands across the surface, imparting thrust to the lightcraft. If the time of the pulse is short such that the plasma remains close to the momentum surface, momentum is imparted and thrust can be generated.

There have been a number of attempts to analyse and model the physics behind the generation of thrust by the plasma. [3, 7] have developed a simplified parametric model of the flow field produced by a high energy laser beam incident onto a non-ablative surface.

The flow field present due to the laser supported detonation process can be generally classified into two different regimes dependant on laser intensity, the Laser Supported Detonation (LSD) wave and the Laser Supported Combustion (LSC) wave. The LSC wave has been observed at flux levels in excess of $3 \times 10^4 W/cm^2$, while the LSD wave is apparent in flux levels around $10^7 W/cm^2$. Figure 2 illustrates the flow regimes present for both types of laser supported waves. A transition flow field often occurs between these intensities.

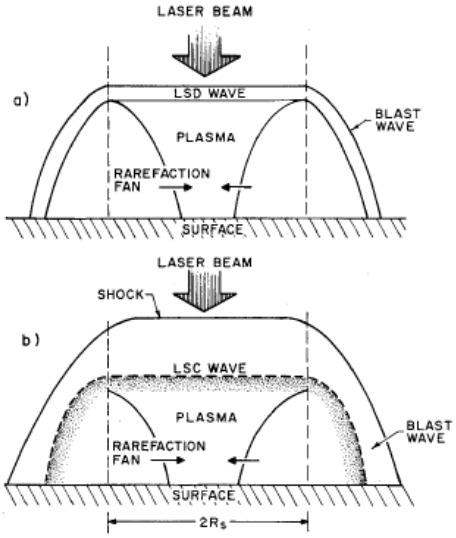


Fig. 2: Schematics of flow regimes: a) LSD wave; b) LSC wave. [7]

The surface pressure time histories for the laser supported detonation wave are described in terms of two characteristic time scales, as shown in the $x-t$ diagrams in Figure 3. The more important time scale in describing the process is the radial relaxation time, which represents the time taken for the rare-fraction wave from the detonation to reach the spot centre. This is given by

$$\tau_{2D} = R_{LSD}/C_{LSD} \tag{3}$$

where R_s is the laser spot radius, and C_{LSD} is the sound speed in the hot plasma. For a laser supported detonation wave, the sound speed is taken as half of the blast wave velocity[3, 7].

The initial surface pressure conditions, based on complete absorption of the laser flux into the detonation front is given by

$$P_{LSD} = \left[\frac{\gamma+1}{2\gamma}\right]^{\frac{2\gamma}{\gamma-1}} \frac{\rho_0 V_{LSD}^2}{(\gamma+1)} \tag{4}$$

where

$$V_{LSD} = \left[2(\gamma^2-1)\frac{I}{\rho_0}\right]^{\frac{1}{3}} \tag{5}$$

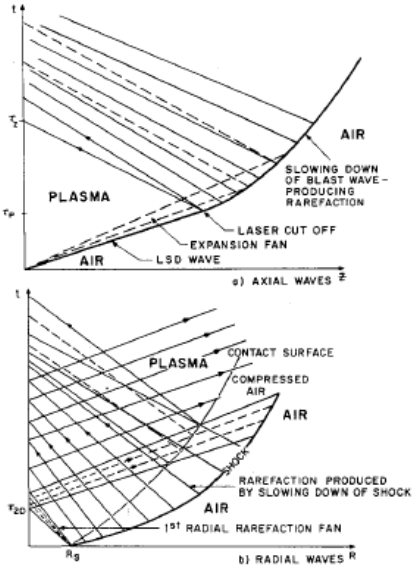


Fig. 3: $x-t$ diagrams of constant flux LSD [7]

Here V_{LSD} is the initial laser supported detonation (LSD) wave velocity in ms^{-1}, and I is the laser intensity in Wm^{-2}, ρ_0 is the density of the working gas in kgm^{-3} and γ is the ratio of specific heats of the working gas. These equations are used to establish the flow and thermodynamic properties of a blast wave that has evolved to cylindrical geometry adjacent a flat plate. The solution to the pressure and velocity profiles are then found employing Sedov's unpowered scaling laws to quantify the time dependant conditions behind the detonation wave for axi-symmetric and isentropic conditions [3].

$$\frac{P}{P_{REF}} = \left(\frac{t}{t_{REF}}\right)^{-1}; \frac{r}{r_{REF}} = \left(\frac{t}{t_{REF}}\right)^{\frac{1}{2}} \tag{6}$$

where P_{REF} can be expressed as the maximum pressure exerted on a flat plate thrust surface by the detonation wave

$$P_{REF} = P_{LSD} \tag{7}$$

The value t_{REF} corresponds to the radial relaxation time and r_{REF} is equivelent to r_{LSD}, the radius when the wave is evolved to a cylindrical geometry. The reference time is therefore given as

$$t_{REF} = \tau_{2D} = \frac{r_{LSD}}{C_{LSD}} = \frac{2r_{LSD}}{V_{LSD}} \tag{8}$$

These first-order estimates neglect real gas effects, such as excitation, dissociation and recombination, however they are suitable for generating initial estimates of the behaviour of the laser induced detonation wave. A simple thermodynamic relation for the expansion of

the plasma is then assumed, and using an equation of state and compressibility factor, Z, the temperature of the gas behind the front can be calculated.

$$T = \frac{C_{LSD}^2}{\gamma Z \frac{R_U}{M}}$$

where M is the molecular weight and $R_u$is the universal gas constant.

The laser intensity is dependent on the design of the ground based laser, and published values vary from anywhere between 1x10$^{10}W/m^2$[7] and 5x10$^{11}W/m^2$[3]. A laser intensity of 1x10$^{11}W/m^2$ has been used in this study, as this ensures the laser supported detonation (LSD) wave is present [7].

The conditions behind the wave front are considered to be constant during the formative growth and plasma development stages. Once the wave is cylindrical, the Sedov scaling laws apply and the front properties decay accordingly. The thrust generated by each individual pulse is obtained by integrating the pressure of the blast wave over the lightcraft afterbody as it expands. Figure 4 shows a typical laser supported detonation wave profile.

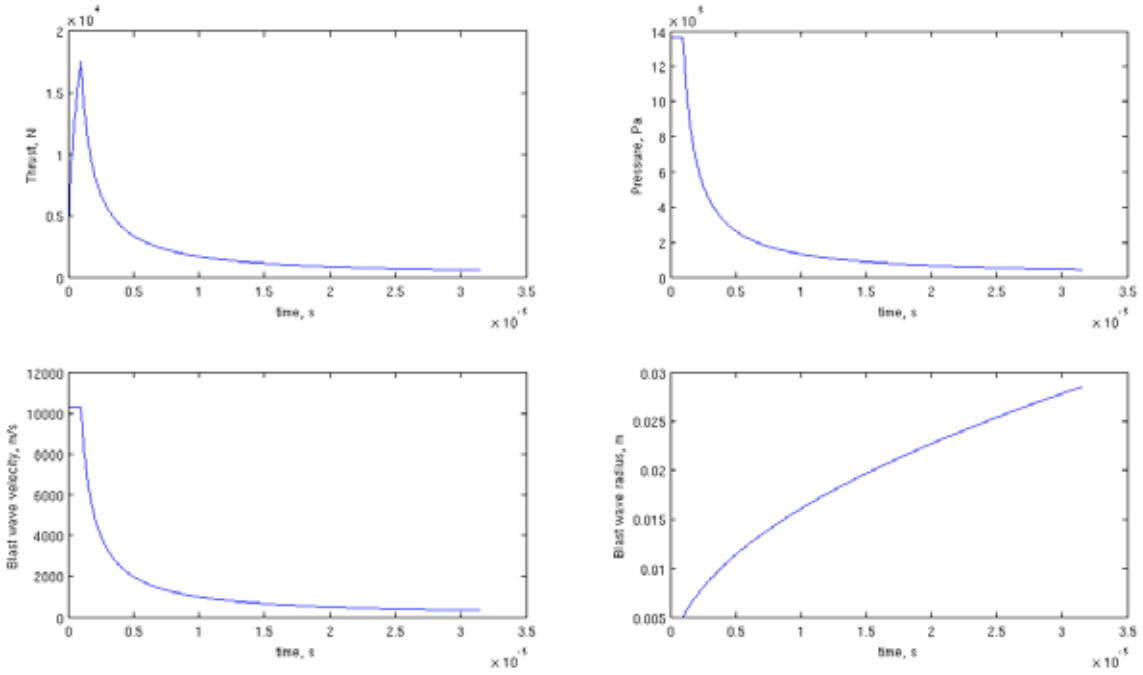


Fig. 4: Typical blast wave properties

The thrust calculation assumes that the pressure behind the wave front does not relax in the radial direction, and hence the pressure behind the front is the same as the decaying pressure of the front. This results in an over approximation of the thrust generated by the laser pulse, however due to the rapid decay of the front the thrust calculated will be in the same order of magnitude as the actual thrust.

Inlet design methodology

The stream traced inlet design methodology is a technique used to design hypersonic inlets. This technique is typically applied to conventionally fuelled scramjet engines [8], but has been modified to suit the requirements of lightcraft. With the stream traced inlet design methodology, specific desired inlet design conditions in the engine can be stipulated, and an inlet geometry generated to suit. The stream traced methodology can be applied, in its simplest form, from two dimensional planar inlets (such as the work of [9]) to complicated three dimensional geometries [10]. This design methodology has been chosen for the lightcraft design due to its performance at off-design conditions[10]. To maintain the lightcraft performance at a range of flight Mach numbers, altitudes and vehicle angles of attack, a fixed geometry inlet that does not un-start is required. It must also be ensured that if un-start does occur, the inlet can be successfully self-starting.

In the stream traced inlet design methodology, the designer begins by defining the desired conditions they wish to achieve in the isolator of the lightcraft. When using the methodology for a scramjet engine, the designer will typically choose a desired pressure within the combustor at a certain inlet entrance Mach number. With the lightcraft design it is required that air (the working gas) be delivered to the laser induced detonation process at specific densities, as this determines the generated thrust. The incoming free stream air is required to be compressed adequatley to achieve the desired results. As the design conditions for the lightcraft are set, the flow properties entering the generating flow field are that of the compressed air due to the conical forebody of the lightcraft performing the initial compression. Once these design parameters have been established, a generating flow field is created that performs the required compression of the inlet flow. For three dimensional inlet designs (including axi-symmetric inlets) inward turning, internal axi-symmetric flow fields with conical compression following a free-standing conical shock at the flow field 'throat' created by the flow field inlet lip are used due to the inherent isentropic compression[11]. The throat shock results in uniform flow that is parallel to the free stream flow.

The inlet geometry is then created by choosing any stream surface as the solid inlet wall. Flow field stream lines are replaced by a solid, inviscid inlet geometry. The desired generating shape is defined either upstream of the compression shock or at the end of the compression field. The streamlines that pass through this defined capture area are then followed downstream to the end of the compression field (or back towards the leading edge shock), defining the inlet shape as illustrated by Figure 5. This method creates a 'swept' surface, reducing heat transfer issues as well as increasing self starting properties[10].

The flow field used to generate the lightcraft inlets is shown in Figure 6. The free-standing shock, and resulting flow cancellation, can be seen in the generating flow field. The streamlines that produced the required inlet cowl shock cancellation at the expansion into the isolator were then used to form the inviscid inlet geometry. A viscous correction was applied to allow for the growing boundary layer within the inlet. The resulting inlet geometry can be seen in Figure 7.

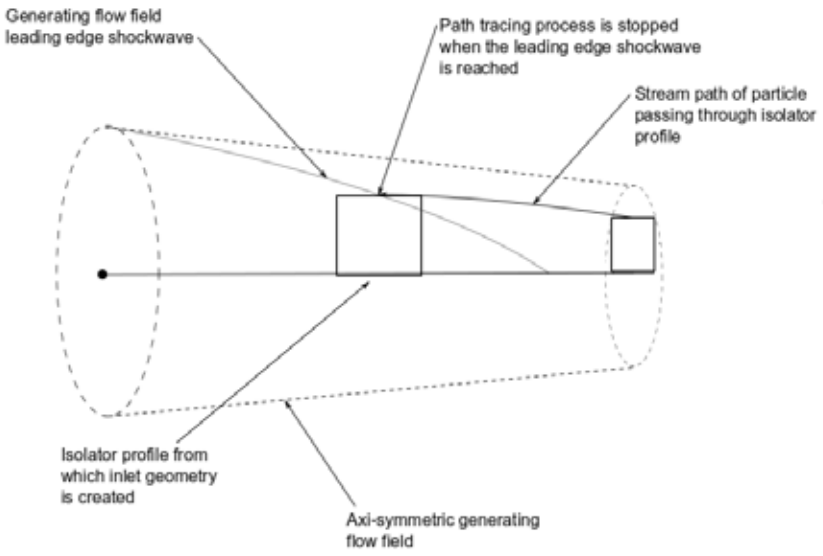


Fig. 5: Tracing of streamlines through generating flow field

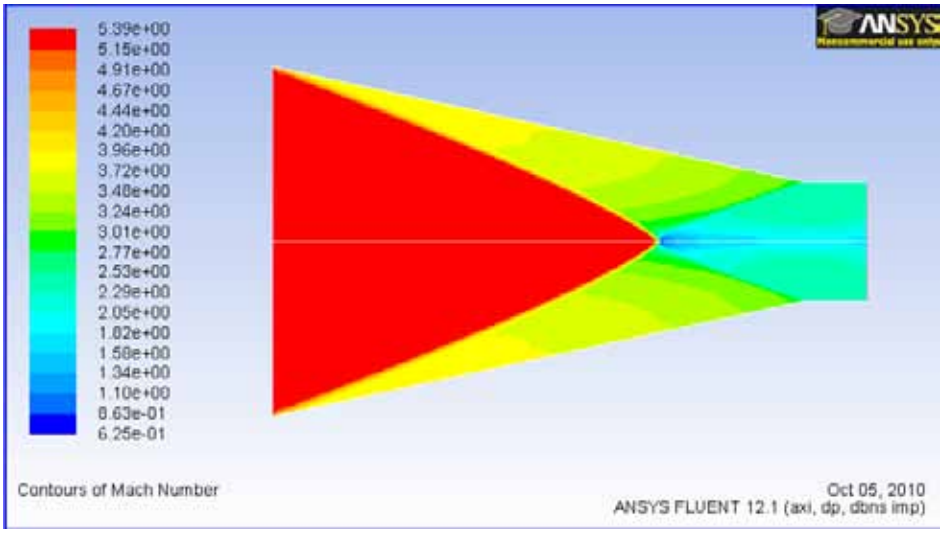


Fig. 6: Generating flow field

Results

A computation fluid dynamics simulation has been performed on the inlet with the detonation wave conditions. The initial values obtained from the analytical solution described in the previous section on laser supported detonation theory, listed in Table 1, are employed in the CFD analysis. In addition, a total time for the laser pulse cycle of 3×10^{-5} seconds was used as the baseline for the evaluation of the inlets ability to restart before the following laser pulse. The performance of the system increases with decreasing pulse cycle duration, with diminishing returns as the pulse cycle reaches the pulse duration (i.e. a continuously working laser) [4]. A range of values for pulse cycle length have been employed in previous studies, ranging anywhere from a continuous energy deposition, up to a cycle duration of 0.1s for subsonic flight tests[12].The value used in the final design will be a compromise between laser capability, inlet performance and thrust generation.

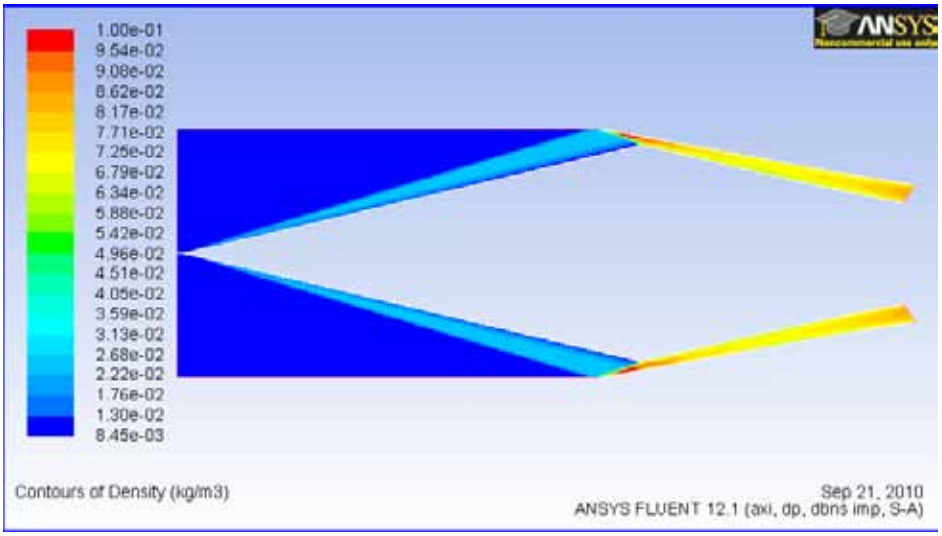


Fig. 7: Final stream traced inlet geometry

Parameter	Value
Radius	5mm
Pressure	1,364,000Pa
Velocity	10,320m/s
Temperature	18,000K

Table 1: Laser induced detonation wave initial conditions

A two-dimensional axi-symmetric turbulent transient simulation was performed in the commercial computational fluid dynamics code Fluent to record the time dependent flow history. The initial flow field shown in Figure 7 was employed to represent the inlet flow field present during cruise flight conditions at zero degrees angle of attack, Mach number of 8 and no laser induced detonation. The values listed in Table 1 were then patched into a cell zone of 5mm, representing the completely cylindrical laser induced detonation wave. This is shown in Figure 8.

Figure 9 shows the transient progression of the detonation wave as it expands and relaxes over the lightcraft afterbody. The detonation wave front can be seen to expand over the lightcraft afterbody as the solution progresses. The detonation wave front can also be seen to move up into the inlet isolator, where it remains until it is exhausted by the inlet flow. The detonation wave remains in the isolator for a period greater than the laser pulse cycle, preventing the working gas from being refreshed for the successive pulses. This is a significant issue for the laser induced detonation process design, as the efficiency and performance will be decreased.

The blockage the laser induced detonation wave creates is further illustrated in Figure 10. The contours show the normal wave where the flow is decelerated to subsonic speeds, before being expanded at the nozzle.

Figure 11 shows the eventual extinction of the laser induced detonation wave, and the subsequent restart of the lightcraft inlet. Although the time scale of the inlet restart is greater

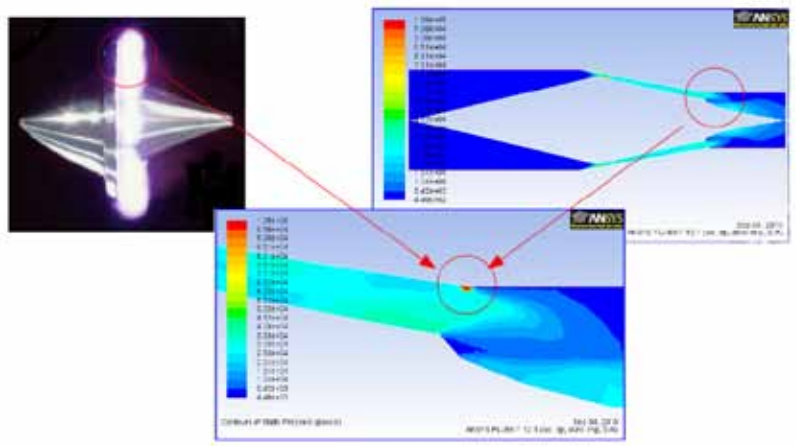


Fig. 8: Laser induced detonation wave simulation initial setup

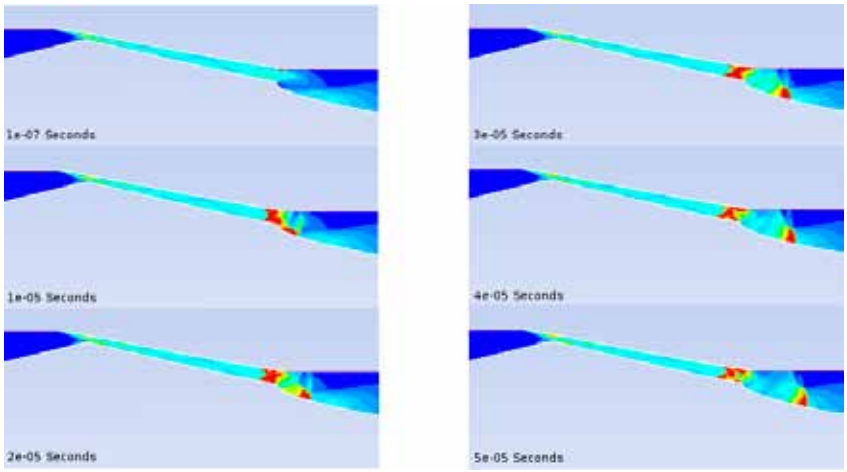


Fig. 9: Contours of pressure for detonation wave evolution

than the laser pulse cycle, it is a promising result that the inlet is able to restart between laser pulses.

Conclusion

From the transient simulation it is clear that great care is required to ensure that the working gas is refreshed before subsequent laser pulses occur in the detonation cycle. This presents a significant design issue that needs to be addressed for lightcraft. If the working gas is not sufficiently refreshed between laser pulses, the inlet performance will suffer and the required thrust will be not be generated. It is apparent, however, that the inlet design does restart given sufficient time. Satisfactory performance can be obtained by altering the laser induced detonation propulsion system design parameters. Factors such as time between laser pulses, inlet flow field characteristics and expansion nozzle geometry all play a part in the performance of the laser propulsion system and an optimised solution is required. Altering an individual aspect of the propulsion system will affect all components of the highly non-linear

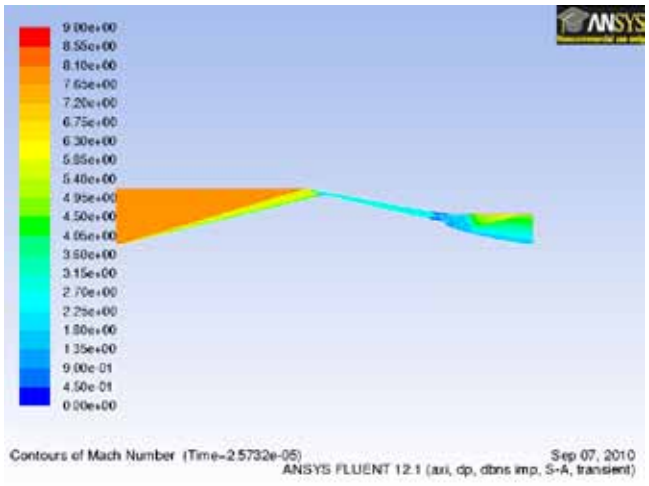


Fig. 10: Inlet isolator normal wave

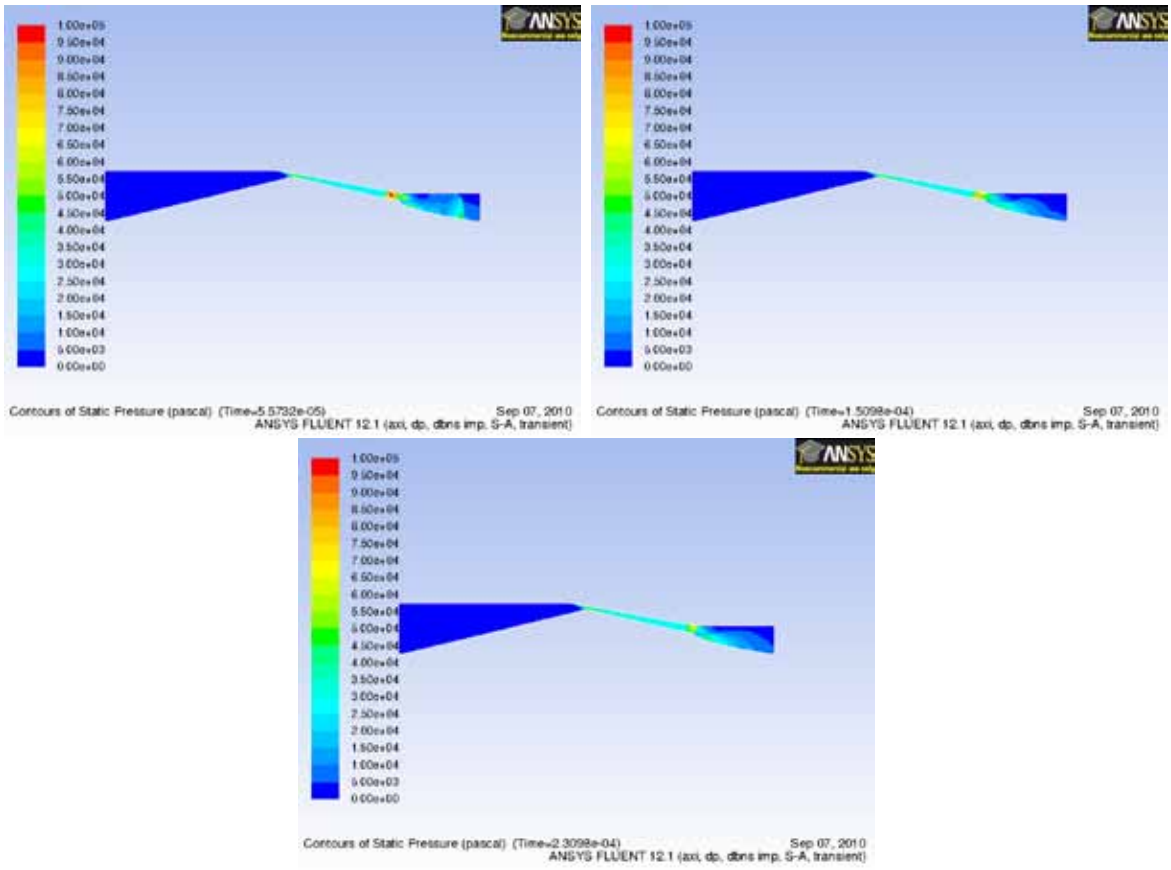


Fig. 11: Contours of pressure for detonation wave extinction and inlet restart

system, and hence a much more detailed performance study is required to deliver the optimal lightcraft propulsion system configuration.

Acknowledgments

The author would like to thank the Asian Office for Aerospace Research and Development for its continued support of this research.

References

[1] Kantrowitz, A., "Propulsion to Orbit by Ground Based Lasers", *Journal of Astronautics and Aeronautics*, Vol 10, pg74-76, May 1972.

[2] Mead Jr, F.B. and Davis, E.W., "Review of Laser Lightcraft Propulsion System", *5th International Symposium on Beamed Energy Propulsion*, November 2007.

[3] Feikema, D., "Analysis of the Laser Propelled Lightcraft Vehicle", *31st Plasmadynamics and Lasers Conference*, American Institute of Aeronautics and Astronautics, June 2000.

[4] Simons, G.A. and Pirri, A.N., "The Fluid Mechanics of Pulsed Laser Propulsion", *American Institute of Aeronautics and Astronautics*, Vol 15, pg835-842, June 1977.

[5] Morgan, C.G, "Laser Induced Breakdown of Gases", *Reports on Progress in Physics*, Vol. 38, No.5, pg621-666, 1975.

[6] Salvador, I.I., "Static and Hypersonic Experimental Analysis of Impulse Generation in Air-breathing Laser-thermal Propulsion", *Master's Thesis*, Rensselaer Polytechnic Institute, 2010.

[7] Ballantyne, A., Reilly, J.P. and Woodroffe, J.A., "Modelling of Momentum Transfer to a Surface by Laser-supported Absorption Waves", *American Institute of Aeronautics and Astronautics*, Vol. 17, pg1098-1105, October 1979.

[8] Matthews, A.J. and Jones, T.V., "Design and Test of a Modular Waverider Hypersonic Intake", *Journal of Propulsion and Power*, American Institute of Aeronautics and Astronautics, Vol. 22, pg913-920, 2006.

[9] Nonweiler, T., "Delta Wing Shape Amenable to Exact Shockwave Theory", *Journal of the Royal Aeronautics Society*, Vol. 67, 1963.

[10] Smart, M.K., "Design of Three-dimensional Hypersonic Inlets with Rectangular-to-elliptical Shape Transition", *Journal of Propulsion and Power*, Vol 15, pg408-416, 1999.

[11] Szpiro, E.J. and Molder, S., "Busemann Inlet for Hypersonic Speeds", *Journal of Spacecraft*, Vol. 3, pg1303-1304, 1966.

[12] Myrabo, L.N., Messitt, D.G. and Mead Jr., F.B., "Ground and Flight Tests of a Laser Propelled Vehicle", *American Institute of Aeronautics and Astronautics*, 1998.

Simulation of Hypervelocity Scramjet Combustion with Oxygen Enrichment

D. Petty [1], M. K. Smart [1], V. Wheatley [1] and S. A. Razzaqi [1]

[1] *Centre for Hypersonics, The University of Queensland, Brisbane, Queensland, 4072, Australia*

Summary: A numerical investigation of oxygen enriched combustion within a hypervelocity scramjet was performed using two-dimensional Reynolds-averaged Navier-Stokes simulations. The simulations modelled previous oxygen enrichment experiments in a simplified scramjet flow-path. It is shown that oxygen enrichment significantly affects the mixing and combustion characteristics within a scramjet combustor. Simulation results indicate that combustion efficiency of the scramjet improved beyond the amount expected from the O_2 premixed with the fuel reacting with the stoichiometric quantity of H_2.

Keywords: hypervelocity, scramjet, oxygen enrichment, air-breathing.

Introduction

Air-breathing propulsion systems, such as the scramjet engine, with access-to-space capability have attracted considerable interest due to the current dependency on multi-stage rockets with low specific impulse. A three-stage rocket-scramjet-rocket system has been proposed to power a launch vehicle into a low Earth orbit [1]. The launch vehicle will travel at near-constant dynamic pressure during the operation of the scramjet. Results from the study showed that such a vehicle would struggle to maintain net thrust above Mach 12. This restricts the usefulness of scramjet engines for space access.

The difficulty of scramjet engines maintaining net thrust at higher altitudes is contributed by both air (specifically O_2) mass capture and residence time of O_2 within the combustor decreasing. These effects can be illustrated by performing a simple analysis. Air density of the atmosphere decreases by three orders of magnitude when travelling from the Earth's surface to an altitude of 47km [2]. Species mass fractions remain effectively constant within these altitudes so the mass of O_2 per unit volume of air decreases linearly with air density. Dynamic pressure variations are minimal in order to generate sufficient lift whilst avoiding excessive drag and heating loads [3]. Using the definition of dynamic pressure (q) and the one-dimensional continuity equation, the mass flow capture ($\dot{m}_C$) of a scramjet engine can be defined as:

$$\dot{m}_C = 2\frac{q_\infty A_C}{v_\infty} \tag{1}$$

This relationship shows that for trajectories which traverse along near constant dynamic pressure paths, the mass flow capture of the engine is inversely related to the flight speed (v_∞). Variations in capture area (A_C) are trivial in comparison to the increase in flight speed as the vehicle accelerates so mass capture must decrease with increasing altitude. The adverse effect of diminishing O_2 mass capture on scramjet performance is further compounded by the decreasing residence time of flow within the scramjet engine as flight speeds increase. It can be concluded from these factors that achieving high combustion efficiency is paramount in

hypervelocity scramjet engines. It is critical that effective mixing of air and fuel at the molecular level to a near stoichiometric mixture of O_2 and fuel occurs [3].

Oxygen enrichment is a technique aimed at supplementing the diminishing amount of O_2 capture at altitudes and speeds where scramjet shutoff is predicted to occur. The addition of O_2 has two additional effects at any altitude: the injected flow shifts towards a stoichiometric mixture of fuel and oxidizer as well as changing the fluid properties of the injected flow [4]. Pike [5] showed for a restricted class of scramjets that the addition of O_2 to the injected fuel decreased launch mass of a vehicle whilst increasing the payload mass.

Shock tunnel experiments simulating hypervelocity conditions were conducted by Razzaqi & Smart [6] to investigate the effectiveness of oxygen enrichment within a simplified scramjet engine. Oxygen enrichment was found to be effective at increasing the thrust of the scramjet, particularly at high altitudes. However, as only pressure transducer point measurements were taken, the details of the flow physics that led to the improved performance were not presented.

This paper attempts to reveal the details of the flow physics by carrying out a numerical investigation of Razzaqi and Smart's experiments. This report is structured as follows: first we present an overview of the Razzaqi and Smart experiment [6], then the numerical methodology employed for this study is described and finally, the results obtained are presented, focusing on combustion efficiency comparisons.

Oxygen Enrichment Experiments

The Razzaqi & Smart experiments were performed in the T4 free piston reflected shock tunnel at The University of Queensland [7]. The experiment utilized a contoured axisymmetric nozzle to produce two effective flight altitude conditions: standard altitude with M_∞ = 12.3 and q_∞ = 40.2 kPa and high altitude with M_∞ = 11.4 and q_∞ = 9.8 kPa. The numerical simulations performed for this study focused specifically on the standard altitude conditions. Two high range PCB pressure transducers were mounted in the stagnation region to record the nozzle-supply pressure [6]. A summary of the selected experimental shock tunnel conditions for simulation is detailed in Table 1.

Table 1: Summary of selected shock tunnel test conditions

Shot	Inflow	Composition	Temp (K)	Pres. (kPa)	Mach No. (-/-)	Mass Flow (kg/sec)
9941	Nozzle Exit	Air	988	6.3	5.7	N/A
	Injector	Off	N/A	N/A	N/A	N/A
9952	Nozzle Exit	N_2	958	5.9	5.9	N/A
	Injector	H_2	300 (t)	300 (t)	~ 1	0.0193
9953	Nozzle Exit	Air	986	6.4	5.7	N/A
	Injector	H_2 (ϕ=0.84)	300 (t)	301 (t)	~ 1	0.0194
9944	Nozzle Exit	Air	950	6.11	5.7	N/A
	Injector	H_2 (ϕ=0.81) + Oxy (*EP*~15%)	300 (t)	471 (t)	~ 1	0.0183 + 0.0272

Sequentially, these shots correspond to air only (no fuel injection), mixing only (fuel injected into N_2), combustion, and combustion with oxygen enrichment. The fuelling conditions of both simulation and experiment are described using the standard definition of fuel equivalence ratio (ϕ) for hydrogen fuel:

$$\phi = \frac{1}{2}\frac{M_{O_2}}{M_{H_2}}\frac{\dot{m}_{H_2,IN}}{0.23\dot{m}_C} \tag{2}$$

Where M_{O_2} is the molar mass of O_2, M_{H_2} is the molar mass of H_2, $\dot{m}_{H_2,IN}$ is the mass flow rate of injected H_2, $\dot{m}_C$ is the mass flow rate of ingested O_2.

The oxygen enrichment shot employed an additional flow (Oxy) of 75/25 (% by wt) O_2/N_2 mixture which was premixed with the fuel flow before injection. The parameter used to describe the amount oxygen enrichment is equivalent to the Enrichment Percentage (*EP*) used by Razzaqi & Smart [6]. This parameter is defined as the percentage of injected H_2 which would be consumed in a stoichiometric reaction with the injected enrichment oxygen:

$$EP = 2\frac{M_{H_2}}{M_{O_2}}\frac{\dot{m}_{O_2,IN}}{\dot{m}_{H_2,IN}} \times 100\% \tag{3}$$

Where $\dot{m}_{O_2,IN}$ is the mass flow rate of injected O_2. The objective of this set of experiments was to investigate the effect of varying EP whilst holding $\dot{m}_{H_2,IN}$ and $\dot{m}_C$ relatively constant.

The experimental model consists of three primary features: a two-dimensional compression wedge at an angle of 8° to the flow, a constant area rectangular combustor (duct) that has an aspect ratio of *l*/*h* = 18.4 and an expansion surface inclined at an angle 11° to the flow direction

The duct has a height of 47 mm and a width of 100 mm. An injector strut with a height of 7 mm injects fuel through a 1 mm × 100 mm slot located on the backward face. A cross-sectional schematic of the experimental test section is shown in the Fig. 1.

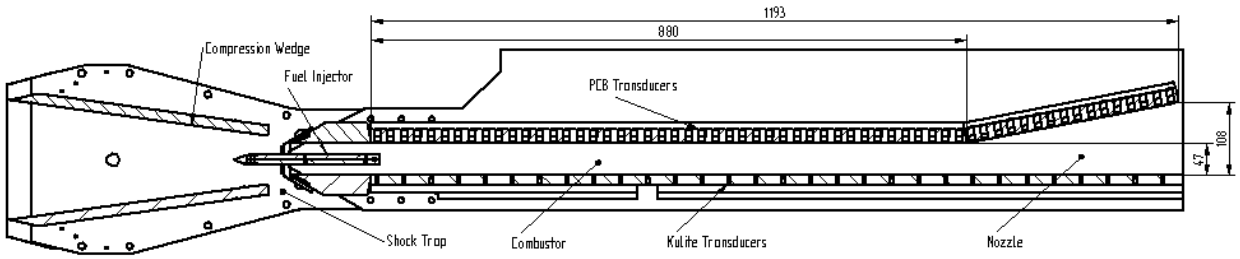


Fig. 1: Experimental model schematic [6]

The findings of the experimental campaign of [6] validated the previous theoretical findings, showing that specific thrust linearly increases with equivalence ratio at a particular oxygen enrichment percentage.

Determination of combustion efficiency from the experimental data required the use of a quasi one-dimensional cycle analysis method [8]. The analysis assumes a thermally perfect mixture of gases in thermal equilibrium. Skin friction drag and heat transfer across the duct walls were approximated using the Reynolds analogy. A combustion efficiency curve, taken from Heiser and Pratt [3], was used to model the effective mixing and kinetics of combustion

Employing this cycle analysis to predict pressure values along the duct, the combustion efficiency curve was manipulated to best match the experimental pressure measurements taken along the lower surface of the duct (Kulite transducers). Once pressure values were matched, the derived combustion efficiency curve was considered representative of the true combustion within the experimental duct, as shown in *Fig. 2*.

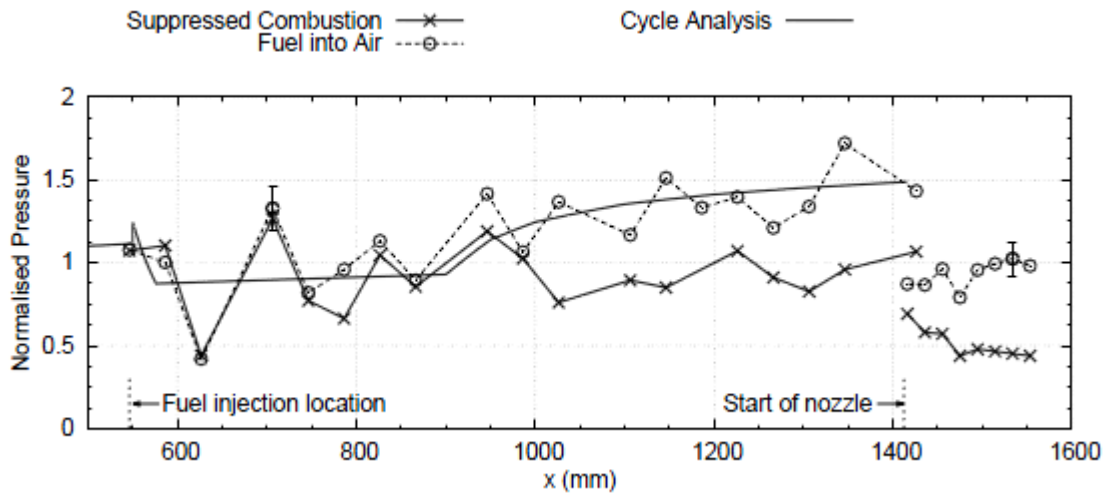


Fig. 2: Typical cycle predicted and experimental pressure distributions [8]

Using this method, the combustion efficiencies for shots 9953 and 9944 were found to be 45% and 60% respectively. It was concluded that combustion efficiency increased by an amount equivalent to the enrichment percentage.

Numerical Methodology

NASA's viscous upwind algorithm for complex flow analysis (VULCAN) version 6.1.0 [9] was the numerical solver used in this investigation. Vulcan can simulate two and three dimensional flows on multi-block structured grids by solving cell-centred integral forms of the Reynolds-averaged Navier-Stokes (RANS) equations [10].

The Diagonalized Approximate Factorization (DAF) temporal advancement scheme was used to achieve steady state solution convergence with a Courant-Friedrich-Lewy (CFL) regime ranging from 0.1 to 3.0.

The fluid is treated as a thermally prefect mixture of gas that obeys empirically derived 3-interval 9-coefficient caloric curve fits, for temperatures between 200 and 20,000 K.
The combustion of hydrogen was modelled using the 9 species and 18 finite rate reactions described and validated by Drummond [11].

Turbulence was modelled using Wilcox's k-ω model [12]. This model was shown to reliably simulate mixing layers by Cutler *et al.* [13]. A second order accurate scheme was used for the convective terms in the turbulence equations. Sutherland's law is used to compute the molecular viscosity. Thermal conductivity is computed from viscosity assuming constant Prandtl number. The molecular diffusion coefficient is computed from viscosity assuming constant Schmidt number.

Numerical modelling of the Razzaqi & Smart experiment is relatively straightforward due to the simple geometry of the experimental model. The two-dimensional grids were separated into two regions: the compression wedge and the combined rectangular combustor with expansion surface. The process was performed using Pointwise version 16.03R2 [14].

Uniform inflow conditions, equivalent to the respective shock tunnel exit conditions, were applied at the inflow boundary of the compression wedge region. The test section walls have been modelled using the isothermal (300K) law of the wall boundary conditions (BC) implemented in Vulcan (unless otherwise stated). Injected inflow conditions utilise the generic subsonic inflow BC implemented in VULCAN. First order extrapolated BCs were applied at the outflow boundaries.

Results

All simulations performed were based on the previously mentioned shock tunnel experiments (refer to Table 1) except the nominal 5% oxygen enrichment case. This simulation used the same injected flow conditions as the nominal 15% oxygen enrichment case but replaced the 75/25 (% by wt) O_2/N_2 mixture with air. A summary of the RANS simulations performed and the corresponding Razzaqi and Smart experiments is detailed in Table 2.

Table 2: Summary of RANS simulations of the Razzaqi & Smart experiment

Simulation	Shot	Cells	Turbulence Model	Wall Model	Chemistry Model
Fuel-off, 2D coarse grid	9941	200K	k-ω	Yes	No
Fuel-off, 2D fine grid	9941	686K	k-ω	Yes	No
Fuel-off, 2D fine grid without wall model	9941	686K	k-ω	No	No
3D compression wedge, coarse grid	9941	2.5M	k-ω	Yes	No
Mixing-only	9952	233K	k-ω	Yes	No
Fuel-on	9953	233K	k-ω	Yes	9×18 FR
5% Oxygen enrichment	N/A	233 K	k-ω	Yes	9×18 FR
15% Oxygen enrichment	9944	233 K	k-ω	Yes	9×18 FR

Duct Entrance Flow Conditions

The initial stage of each simulation required determining the flow through the compression wedge into the duct. The flow exiting this plane was used as the inflow conditions for the second grid region (which models the duct and expansion surface).

Compressible two-dimensional flow (oblique shock and Prandtl-Meyer expansion) relations were used to determine theoretical duct entrance conditions for each experimental shot for comparison. A summary of the computed duct entrance conditions is detailed in Table 3.

Table 3: Summary of computed duct entrance conditions

Shot	Method	Temp. (K)	Static Pres. (kPa)	Vel_x (m/sec)	$\dot{m}_C$ (kg/sec)
9941	Theoretical	1877	40.6	3200	0.791
9941	2D RANS	2060	40.8	3070	0.704
9952	Theoretical	1824	39.5	3160	0.782
9952	2D RANS	2000	39.9	3020	0.685
9953	Theoretical	1893	39.9	3330	0.781
9953	2D RANS	2130	42.0	3180	0.701
9944	Theoretical	1874	40.7	3200	0.792
9944	2D RANS	2060	40.8	3100	0.703

Injector Flow Conditions

Injector conditions were calculated using a quasi one-dimensional flow analysis assuming a thermally and calorically perfect gas with the following boundary conditions: a total temperature of 300 K, choked flow through the injector throat and a mass flow rate based on experimentally measured values (refer to Table 1).

Validation

Simulations using a fine grid without a wall function, a fine grid with a wall function and a medium grid with a wall function of the fuel-off case (Shot 9941) were performed to ensure grid convergence. Comparison of the results obtained using these computational grids, based upon pressure values along the duct centreline, showed acceptably minor variation, as shown in Fig. 3.

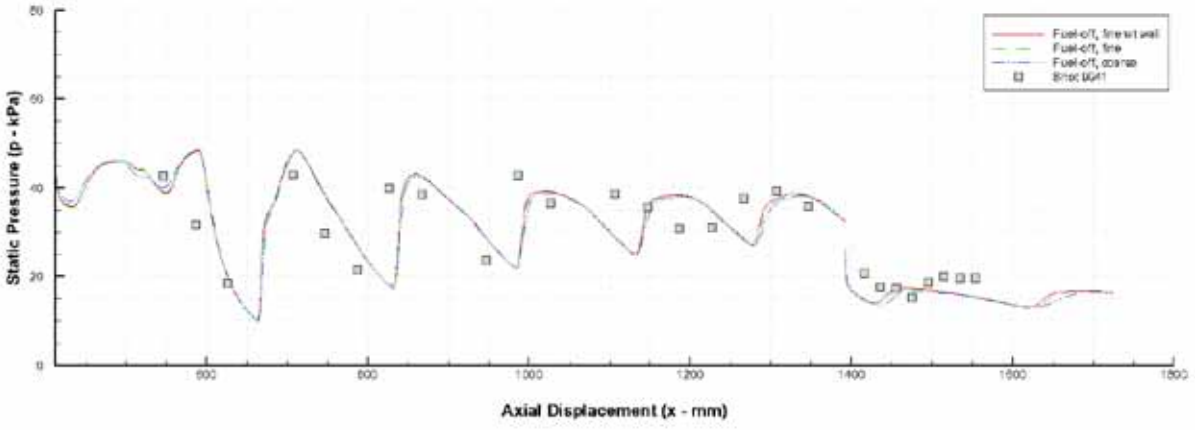


Fig. 3: Air Only CFD and experimental duct static pressure comparison

A comparison between the experimentally measured and simulated static pressures along the duct centreline was performed in order to validate the simulation, as shown in Fig. 3 for the air only case (Shot 9941) and Fig. 4 for the combustion cases (Shot 9953 & 9944). The magnitude of simulated and experimentally measured pressures is in close agreement. However, the difference between simulated and experimental shock reflection locations increases as the flow travels downstream towards the expansion surface. Three dimensional effects may be responsible for these discrepancies, as discussed in a later section.

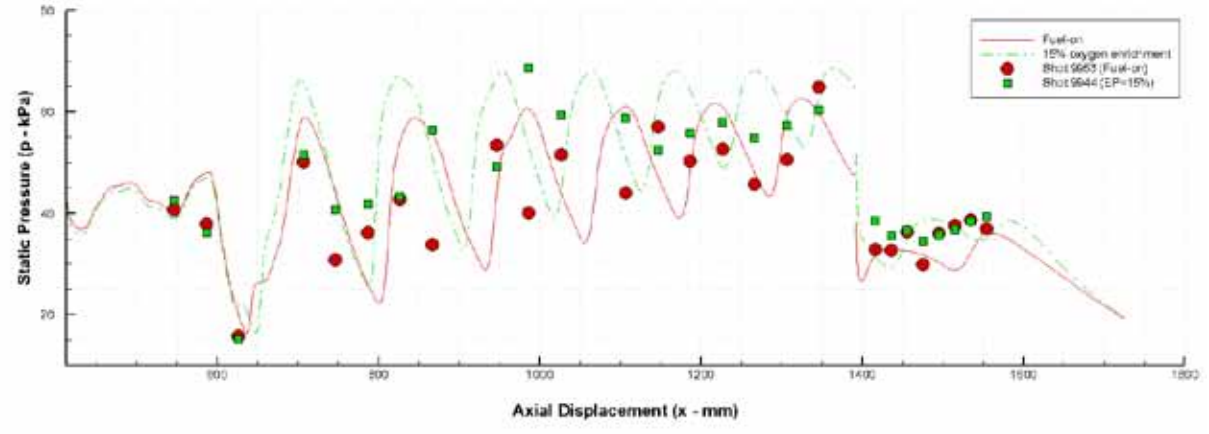


Fig. 4: Fuel into Air CFD and experimental duct static pressure comparison

The simulations predict the experimentally observed pressure rise along the duct for flows which experience combustion, as shown in Fig. 4. Positioning of the shock reflections is more difficult to discern from the experimentally measured pressures since fluctuations are not as clearly evident for the combustion cases.

Recirculation Region

A noteworthy feature of the mean flow through the duct is the recirculation region which forms behind the injection strut, shown in Fig. 5. This region contains low pressure recirculated flow which causes injected flow to be under-expanded just downstream from the point of injection. The injected flow expands until it meets the ingested flow travelling through the duct. In the fuel-on case, a Mach disk then forms in the injected flow causing it to become subsonic. In the nominal 15% oxygen enrichment case, the flow structure is altered: the injected flow is recompressed by a series of oblique shocks rather than by a Mach disk. The strength of the shear layer generated between the injected and ingested flows, which should induce strong mixing between the streams, varies significantly between the cases.

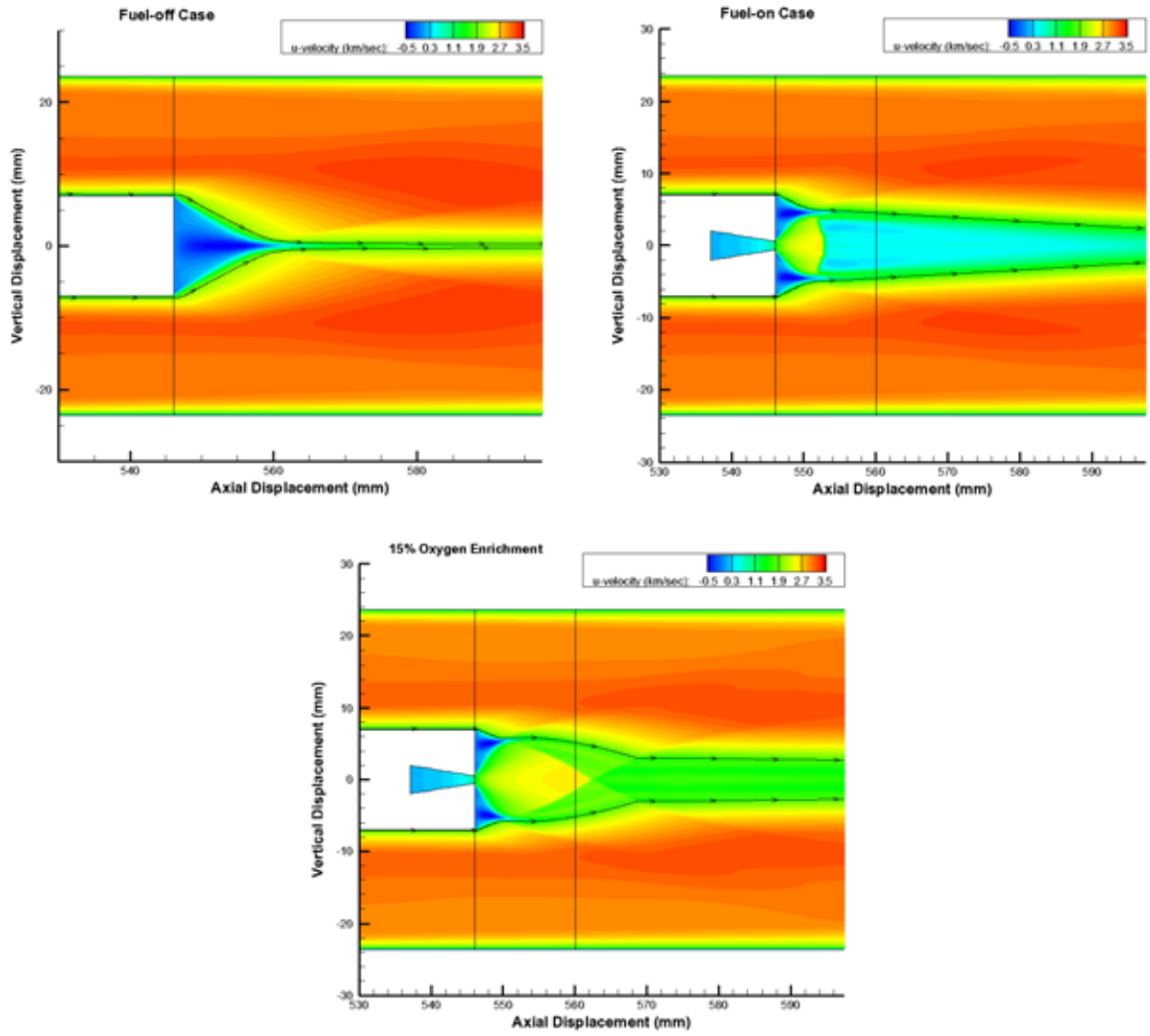


Fig. 5: Axial velocity contours behind the injector strut

Mixing and Combustion Characteristics

Ignition, identified by the net production of OH^* radicals, occurs approximately 40 mm downstream from the point of fuel injection (both for fuel only and oxygen enrichment cases) as can be seen in Fig. 6. The intensity of OH^* production is much higher in the oxygen enrichment case.

Completion of H_2 combustion, identified by the net production (formation) of H_2O, occurs within 1mm downstream of initial liberation of OH^* radicals (both for fuel only and oxygen enrichment cases), shown in Fig. 7. Further, a significantly stronger reaction region occurs at an axial displacement of 760 mm in the oxygen enrichment case.

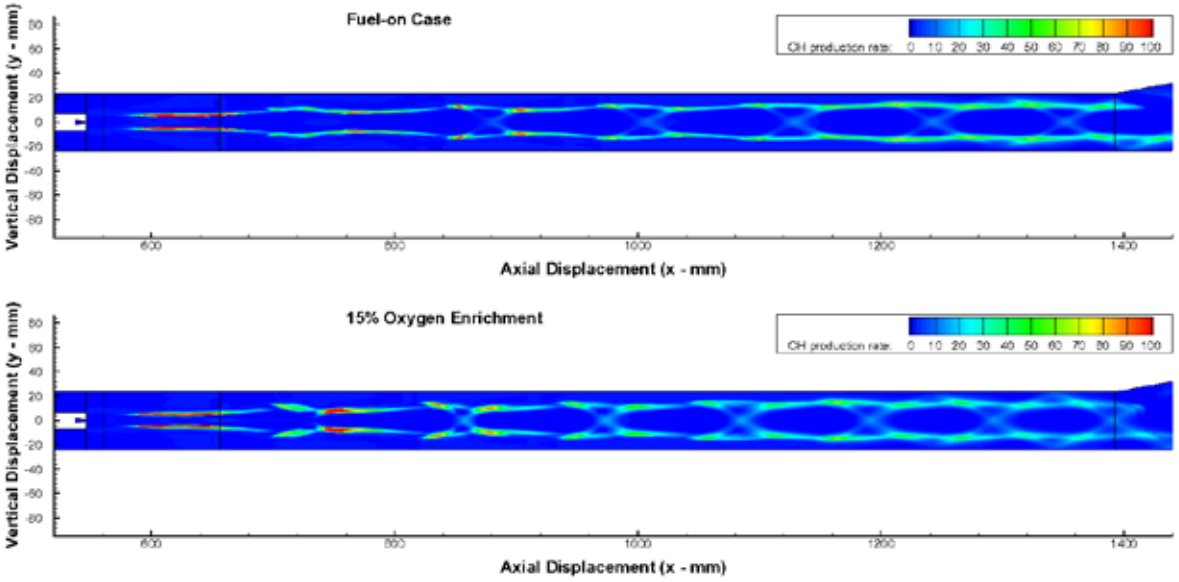


Fig. 6: OH production rate contours along the duct

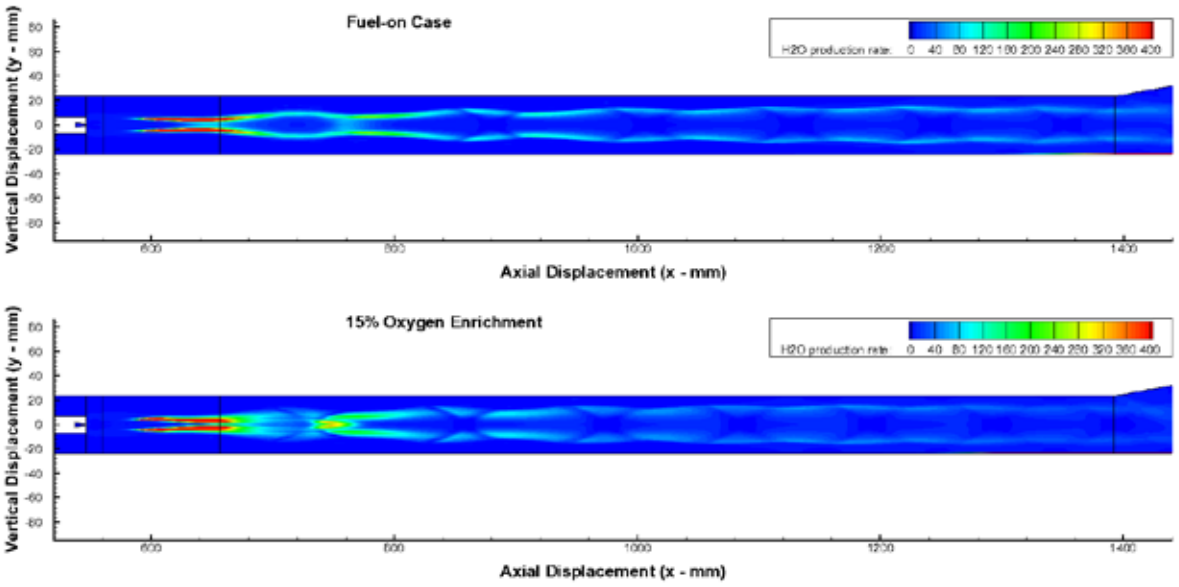


Fig. 7: H_2O production rate contours along the duct

The additional flow characteristics apparent in the oxygen enrichment case noticeably burn a greater amount of the injected H_2, shown in Fig. 8.

The combustion efficiency (η_C) of the combustor is give by the ratio of H_2O mass flow exiting the simulated combustor to the expected H_2O mass flow produced by completely combusting the injected H_2:

$$\eta_C = \frac{M_{H_2}}{M_{H_2O}} \frac{\dot{m}_{H_2O,EXIT}}{\dot{m}_{H_2,IN}} \tag{4}$$

Where M_{H_2O} is the molar mass of water and $\dot{m}_{H_2O,EXIT}$ is the mass flow of water exiting the duct.

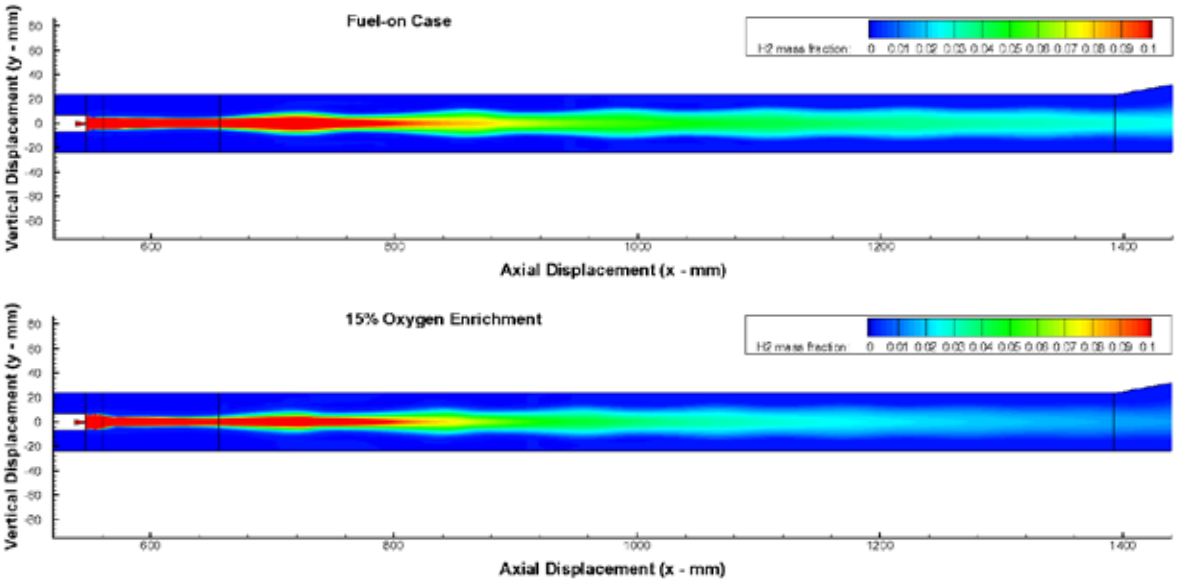


Fig. 8: H_2 mass fraction contours along the duct

For the fuel only case, combustion can only occur between the injected H_2 and the ingested O_2. However, oxygen enrichment cases provide a second source of O_2 which is premixed with the injected H_2. This allows the injected H_2 to combust with both the premixed O_2 and the ingested O_2. For this reason, in the case of oxygen enrichment, combustion efficiency is not a direct measure of the increased combustion strictly caused by the available H_2 combusting with the atmospheric O_2.

It is reasonable to expect the stoichiometric amount of H_2 premixed with the enrichment oxygen will combust completely due to the high combustor temperature. Subsequently, this amount of H_2 cannot react with the ingested atmospheric O_2 residing in the combustor. The efficiency with which the remaining H_2 (after premixed combustion) reacts with the atmospheric O_2 is defined as the compensated combustion efficiency ($\eta_{C,EPC}$):

$$\eta_{C,EPC} = \frac{100\% \times \eta_C - EP}{100\% - EP} \tag{5}$$

A summary of combustion parameters is detailed in Table 4.

Table 4: Summary of Combustion Parameters

Simulation	ϕ CFD	EP CFD (%)	Temp. (K)	η_C Experiment	η_C CFD	$\eta_{C,EPC}$ CFD
Fuel-on	0.84	N/A	2540	45	52.32	N/A
5% Oxygen enrichment	0.81	4.4	2610	N/A	57.98	56.04
15% Oxygen enrichment	0.81	14.1	2730	60	62.72	56.40

All three simulations have a similar equivalence ratio of almost unity. The simulation results show for a premixed fuel percentage of 4.4%, an additional 5.7% of the injected H_2 is burnt in comparison to the fuel only (no premixing) case. This implies that an additional 1.3% of the injected H_2 has reacted with the ingested O_2 due to the secondary effects of oxygen enrichment.

As previously mentioned, the premixed fuel combustion reduces the amount of injected H_2 available to mix and combust with the ingested O_2. The compensated combustion efficiency values show that both oxygen enrichment cases cause an additional 4% of the remaining available H_2 (after premixed combustion) to react with the ingested O_2.

Three-Dimensional Effects

One possible explanation for the mismatch of pressure distributions between simulated and experimental results could be the three-dimensional effects caused by the sidewalls of the experimental model. To investigate the uniformity of the flow entering the duct in the cross-stream direction, a three-dimensional simulation of the compression wedge was performed. Cross-flow velocities of up to 8% of the axial flow velocities were observed in the results of this simulation, shown in Fig. 9. This indicates that three dimensional effects may contribute significantly to the discrepancies between the numerical and experimental pressure distributions along the duct centreline. For example, changes in the axial Mach number within the duct will alter the location and strength of the shock pattern.

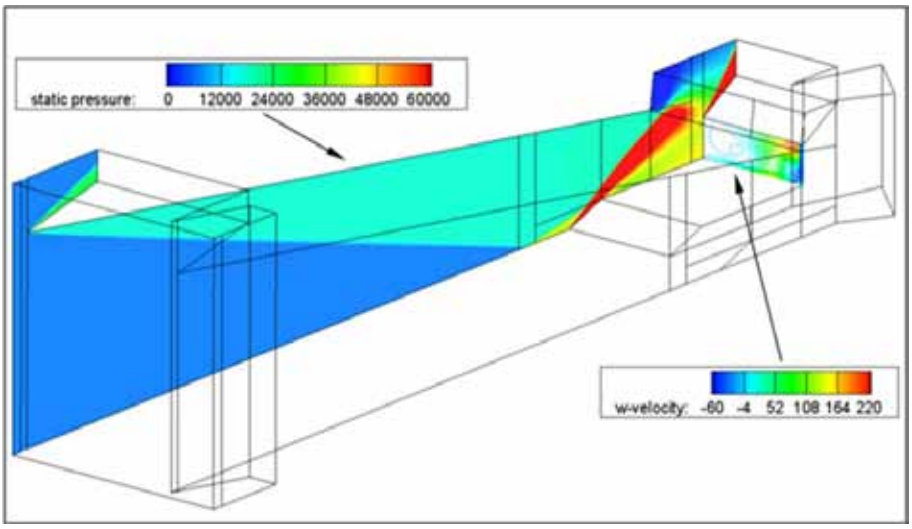


Fig. 9: Static pressure through the wedge and cross-flow velocity entering the duct

Conclusion

A numerical investigation using Reynolds-Averaged Navier-Stokes simulations was presented. This investigation provides new details on the flow physics within an oxygen enriched scramjet combustor. The simulation results show that oxygen enrichment affects the internal flow structure within the particular scramjet combustor. Specifically, oxygen enrichment affects the mixing and combustion characteristics within the combustor. The influence of oxygen enrichment on combustion is most significant in the ignition region. Improvements to combustion efficiency can be made beyond the percentage of premixed and injected O_2. For example, an enrichment percentage of 4.4% produces an increase in combustion efficiency of 5.7%.

References

1. Smart M.K. and Tetlow, M.R., "Orbital Delivery of Small Payloads using Hypersonic Airbreathing Propulsion", *Journal of Spacecraft and Rockets*, Vol. 46, No. 1, 2009, pp. 117-125

2. NASA "U.S. Standard Atmosphere 1976", NASA-TM-X-74335

3. Heiser W.H. and Pratt, D.T., *Hypersonic Airbreathing Propulsion*, AIAA Education Series, AIAA, Washington DC, 1994

4. Rudakov, A.S and Krjutchenko, V.V., "Additional Fuel Component Application for Hydrogen Scramjet Boosting", *SAE Aerospace Atlantic Conference and Exposition*, Dayton, OH, USA, April 1990

5. Pike, J., "The Choice of Propellants: A Similarity Analysis of Scramjet Second Stages", *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*, Vol. 357, 1999, pp. 2357-2378

6. Razzaqi, S.A. and Smart, M.K. "Shock Tunnel Experiments on Oxygen Enrichment in a Hydrogen Fuelled Scramjet", *16th AIAA/DLR/DGLR International Space Planes and Hypersonic Systems and Technologies Conference*, AIAA 2009-7429, Bremen, Germany, Oct. 19-22, 2009

7. Stalker, R. J., Recent Developments with Piston Free Drivers, *17th International Symposium on Shock Waves and Shock Tubes*, 1990, pp 96-107

8. Razzaqi, S.A., private communication, Sep. 2010

9. White, J.A. and Morrison, J.H., "A Pseudo-Temporal Multi-Grid Relaxation Scheme for Solving the Parabolized Navier-Stokes Equations", AIAA Paper 99-3360, 14th AIAA Computational Fluid Dynamics Conference, Jun. 28 – Jul. 1, 1999.

10. NASA, Vulcan version 6.1.0 User Manual, 2009

11. Drummond, J.P., "Numerical Simulation of a Supersonic Chemically Reacting Mixing Layer", NASA TM 4055, 1988

12. Wilcox, D.C., *Turbulence Modeling for CFD*, 3rd edn, DCW Industries, La Cañada, 2006

13. Cutler, A. D., Diskin, G. S., Drummond, J. P., and White, J. A., "Supersonic Coaxial Jet Experiment for Computational Fluid Dynamics Code Validation" *AIAA Journal*, Vol. 44, No. 3, 2006, pp. 585-592.

14. Pointwise Inc., Pointwise version 16.03R2 User Manual, 2010

Physical Insight into Nozzle Flow Behaviour of Axisymmetric Scramjets for Access-to-Space via Design Optimisation

Hideaki Ogawa* and Russell R. Boyce*

* *Centre for Hypersonics, School of Mechanical and Mining Engineering, The University of Queensland, St. Lucia, Brisbane, QLD 4072, Australia*

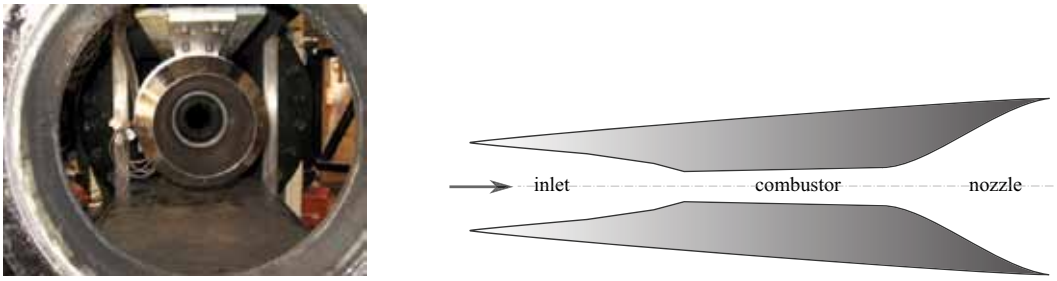
Summary: Scramjet propulsion is a promising hypersonic airbreathing technology for efficient and reliable access-to-space. Internal-compression axisymmetric scramjet (supersonic combustion ramjet) engines are now being explored in the SCRAMSPACE project. The nozzle section plays a key role in producing thrust in scramjet operation. Shape optimisations have been performed for axisymmetric scramjet nozzles to achieve the maximum thrust with various inflow conditions at Mach 8. They have resulted in optimum geometries characterised by similar nozzle length and exit radius for all the tested inflow assumptions, while the major influence of inflow non-uniformity on the optimum contour has been found only in the vicinity of the nozzle entrance. Comparable performance has been obtained in cross-referencing the optimum geometries under different inflow conditions, suggesting detailed geometric difference apart from that in the nozzle length and exit radius exert rather minor impact on the off-design nozzle performance for the conditions considered here.

Keywords: hypersonic airbreathing propulsion, axisymmetric scramjet nozzle, design optimisation, access-to-space

Introduction

Hypersonic air-breathing propulsion offers great potential for reliable, reusable and economical systems for access-to-space. In particular, scramjet (supersonic combustion ramjet) propulsion is a promising technology that can enable efficient and flexible transport systems. The last decade has seen remarkable milestones achieved by various flight experiments including University of Queensland (UQ)'s HyShot II in July 2002 [1, 2], the NASA X-43 vehicles in the Hyper-X program in March and November 2004 [3], and the Boeing X-51A WaveRider in May 2010 [4].

The SCRAMSPACE program is now underway as an international collaboration led by UQ, where internal-compression axisymmetric scramjets (Figure 1 (a)) are being investigated in both ground and flight tests, following the performance demonstrated in shock tunnel experiment [5]. Featuring innovative techniques such as inlet fuel injection and radical-farming shock-induced combustion, the elegantly simple axisymmetric geometry offers numerous advantages in various aspects including aerodynamic and combustion efficiency, thermal and structural management as well as manufacture. Scramjet operation typically relies on the sequential process — capture and compression of hypersonic airflow in the inlet, fuel injection into the air, supersonic combustion in the chamber, and expansion of combustion products for thrust through the nozzle (Figure 1 (b)).



(a) upstream view [5] (b) schematic diagram

Fig. 1: Axisymmetric scramjet engine

The nozzle plays a primary role in producing the thrust to propel the vehicle. The design of axisymmetric scramjet nozzles requires particular care due to complex aerodynamic phenomena and chemical kinetics; the nozzle inflow is highly non-uniform due to the presence of shock-shock and shock wave / boundary layer interactions, which are responsible for the creation of localised hot pockets that are essential for radical-farming shock-induced combustion [6]. Re-association of molecules subsequently occurs in the nozzle as the reacted gas cools in rapid expansion. The fixed geometry makes the nozzle susceptible to performance losses in off-design operation at various velocities and altitudes with / without fuel injection in the course of the SCRAMSPACE mission.

This paper presents the results of nozzle design optimisations for axisymmetric scramjets at an operating Mach number of 8, aiming at the application of hypersonic airbreathing propulsion to efficient access-to-space. Various inflow conditions have been assumed, including uniform and profiled inflows in the presence or absence of fuel, and the flowfields associated with the optimum nozzle geometries have been probed in order to investigate the influential factors and key flow physics.

Approaches

Configurations

Nozzle geometry

The present study is conducted with focus on the nozzle section of an axisymmetric scramjet engine, as seen in Fig. 2. Both internal and external nozzle surfaces are considered here in order to take into account the flow interactions downstream of the trailing edge. The nozzle radius at entrance is fixed at a nominal value 0.03512 m. The inner nozzle surface is represented by an arc defined by the radius r_a and angle θ_a at the entrance followed by a cubic Bézier curve defined by the front vector magnitude λ_n and an inner trailing-edge curve represented by a vector of magnitude λ_{ite} and direction δ_{ite}, which is an angle relative to the straight line that connects both ends of the inner nozzle curve (shown by a dashed line). This geometric representation is adopted particularly because thrust values comparable to those yielded by thrust-optimised nozzles can generally be produced by parabolic nozzles, whose inner surface is represented by a cubic Bézier curve [7]. The position of the trailing edge is dictated by the nozzle length l_n and exit radius r_n. The external nozzle surface is also represented by a cubic Bézier curve defined by vector magnitudes λ_r and λ_{ote} and a relative direction δ_{ote} in a similar manner to the inner curve.

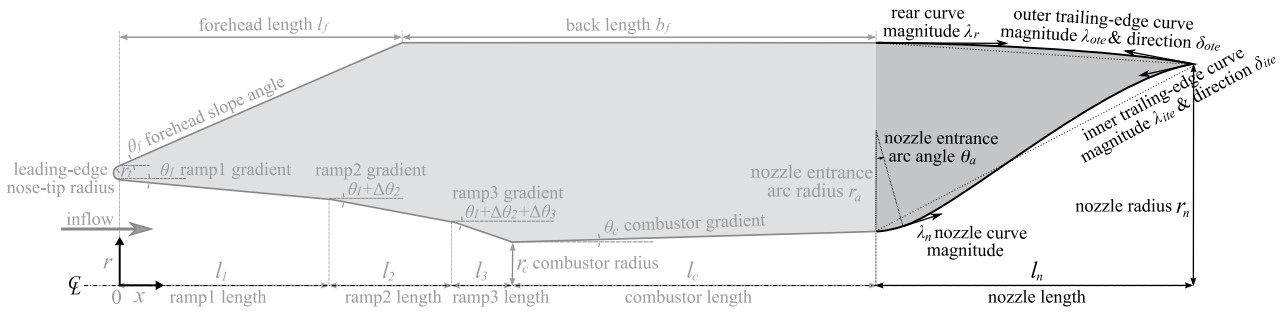


Fig. 2: Parameterisation of the scramjet nozzle geometry

Flow conditions

The freestream condition considered in this study is $M_\infty = 8.0$, $p_\infty = 1197$ Pa, and $T_\infty = 226.5$ K, assuming scramjet operation at an altitude of 30 km on a constant dynamic pressure trajectory of 53.6 kPa. The Reynolds number based on the nozzle entrance radius is $\mathrm{Re}_\infty = 1.06 \times 10^5$.

Computational fluid dynamics

Computational methods

Nozzle flowfields are computed by utilising the state-of-the-art commercial solver CFD++ [8]. An implicit algorithm with second order spatial accuracy is used to solve the Navier-Stokes equations for steady flowfields and convergence is accelerated by the multigrid technique. The nozzle surface is assumed to be an isothermal cold wall of 300K. The boundary layer is assumed to be turbulent and modelled by the two-equation SST $k-\omega$ RANS model [9]. Computations are performed until the energy residual drops to the order of 10^{-5}, based on a convergence dependency study, where all objective functions have been found to vary less than 0.01% at higher orders. Standard air in thermochemical equilibrium state is assumed for the gas in the fuel-off case, whereas the gas composition and finite-rate chemical reactions are represented by Evans and Schexnayder's model [10], which consists of 25 elementary reactions among 12 species including hydrogen-air combustion as well as nitrogen chemistry.

Computational mesh

The computational domain is represented by a two-dimensional structured mesh generated by Glyph scripting within the commercial grid generator Pointwise [11]. The numbers of nodes and cells for the mesh resolutions considered in the present research are tabulated in Table 1[1]. Fig. 3 shows a coarse computational mesh for the baseline geometry with a conical nozzle of a semi-vertex angle of 25°. The minimum grid width is 10^{-5} m at the first grid point off the wall (non-dimensional distance $y^+ = 0.32$ on average) in order to assure good resolution of the viscous sublayer. The coarse resolution level is adopted for the present study, based on a preliminary mesh convergence study for the baseline geometry (results of the mesh sensitivity study are not presented here but this topic is revisited later in the results section).

[1]In the absence of the nozzle entrance arc, the number of nodes along the wall is 10 / 20 / 40 less than those shown for the coarse, fine and superfine mesh, respectively.

Table 1: Mesh resolutions

resolution	along the wall	wall-normal direction	total number of cells
coarse	262	100	25,839
fine	524	200	104,077
superfine	1048	400	417,753

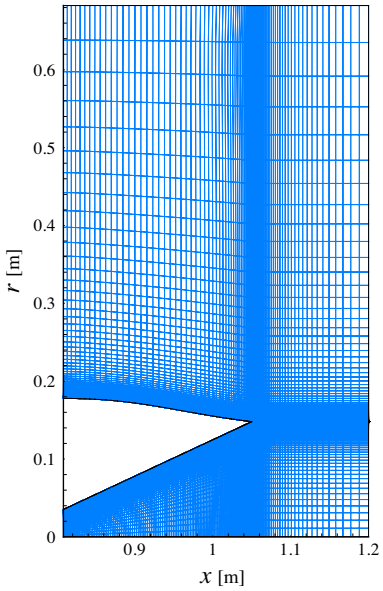


Fig. 3: Computational mesh (baseline)

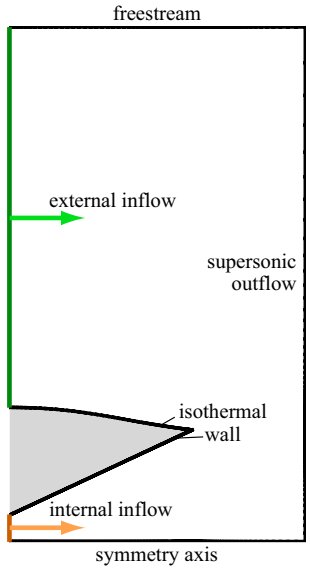


Fig. 4: Boundary conditions

Flow and boundary conditions

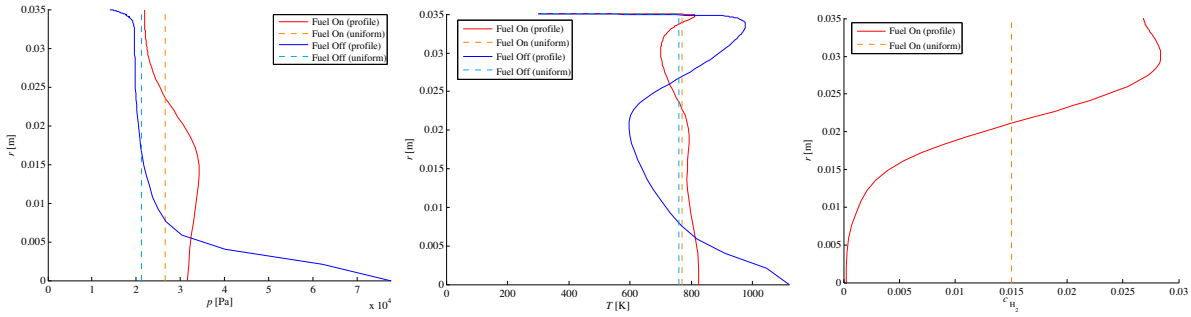
Fig. 4 schematically shows the conditions that are imposed on the boundaries of the computational domain. The surface is assumed to be isothermal walls of 300 K. Imposed at the nozzle entrance are inflow boundary conditions given by either flow profiles or uniform flows, as shown by solid and dashed lines in Fig. 5 (b), respectively. The profiles are obtained by slicing the flowfields calculated in separate axisymmetric CFD runs for the nominal SCRAMSPACE specification[2], except the one with fuel on, which is obtained by circumferentially averaging the sliced flow quantities from a three-dimensional CFD run in the presence of inlet fuel (hydrogen) injection through multiple portholes. The uniform flows are represented by stream-thrust-average [12] values of the profiles. The freestream condition is also applied to the top boundary, while a supersonic outflow condition is imposed on the right hand boundary, which is positioned 0.15m downstream of the trailing edge.

A. Design Optimisation

1) Optimisation algorithms

Design optimisation is performed in an iterative manner as per the optimisation loop shown in Fig. 6, which consists of mesh generation (pre-processing), CFD computation (evaluation), post-processing and optimisation algorithms.

[2] $r_t = 0.0005$ m, $l_1 = 0.1994$ m, $\theta_1 = 5.90°$, $l_2 = 0.069$ m, $\Delta\theta_2 = 3.18°$, $l_3 = 0.040$ m, $\Delta\theta_3 = 3.13°$, $r_c = 0.035$ m, $l_c = 0.5$ m, $\theta_c = 0°$, $r_a = 0$ m, $\theta_a = 0°$, $\lambda_n = 0$ m, $l_n = 0.242$ m, $r_n = 0.148$ m, $\lambda_{ite} = 0$ m, $\delta_{ite} = 0°$, $l_f = 0.280$ m, $\theta_f = 20.0°$, $b_f = 0.528$ m, $\lambda_r = 0.1$ m, $\lambda_{ote} = 0$ m and $\delta_{ote} = 0°$ in Fig.2.



(a) pressure (b) temperature (c) hydrogen mass fraction

Fig. 5: Internal inflow profiles

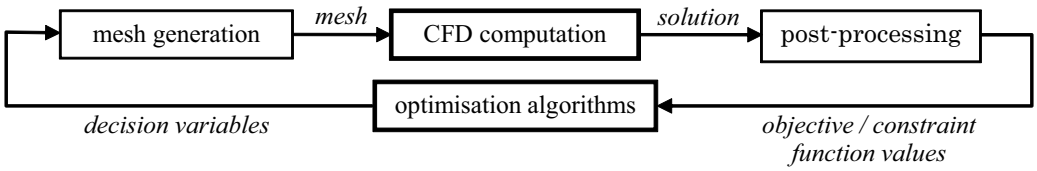


Fig. 6: Optimisation loop

Population-based evolutionary algorithms developed at UNSW@ADFA are employed for optimisation [13, 14]. In particular, use is made of the elitist non-dominated sorting genetic algorithm (NSGA-II) [15]. Optimisation occurs over generations with a population of 32 individuals. A simulated binary crossover and polynomial mutation are used as recombination operators at a given probability (1.0 and 0.1, respectively) with a specified distribution index (10 and 20, respectively). The optimisation process is efficiently assisted by various surrogate models including the response surface models, kriging approximations and radial basis functions. Among these models the one with the least error within a threshold of 10% is adopted to predict the objective and constraint functions in lieu of actual CFD evaluation and all individuals that are estimated to be superior to the present elitists are verified by true CFD evaluation. All members in the population pool are truly evaluated by CFD every 5 generations, when the surrogate models are trained by using 90% of the solutions from the archive, which stores all solutions that are evaluated by true evaluation.

2) Optimisation problem

The nozzle performance is assessed by the sum of the forces acting on the inner and outer nozzle surface: $F \equiv F_{\text{inner}} + F_{\text{outer}}$. This parameter is used as the objective function to be maximised[3] in order to set the goal of the design optimisation. No constraint function is employed for the present study. The 10 design parameters that define the nozzle geometry (Fig. 2) are used as the decision variables, the upper and lower bounds of which are shown later at the top and bottom of the columns in the graph Fig. 8. Design optimisation has been performed with four different assumptions for the gas and internal inflow boundary condition (refer to Fig. 4) as tabulated below:

[3]Note that the sign of F is negative for positive thrust due to the definition of the x axis.

Table 2: Nozzle inflow conditions

fuel (hydrogen)	gas assumption	internal inflow
on	non-equilibrium	profiled
on	non-equilibrium	uniform
off	equilibrium	profiled
off	equilibrium	uniform

Results

Optimisation results

Optimisation progress

The nozzle geometry has been optimised to attain maximum possible thrust for the four inflow assumptions. Fig. 7 shows the progression of the thrust values for each case during optimisation. It is seen that convergence has been achieved as at the 30^{th} generation in all cases.

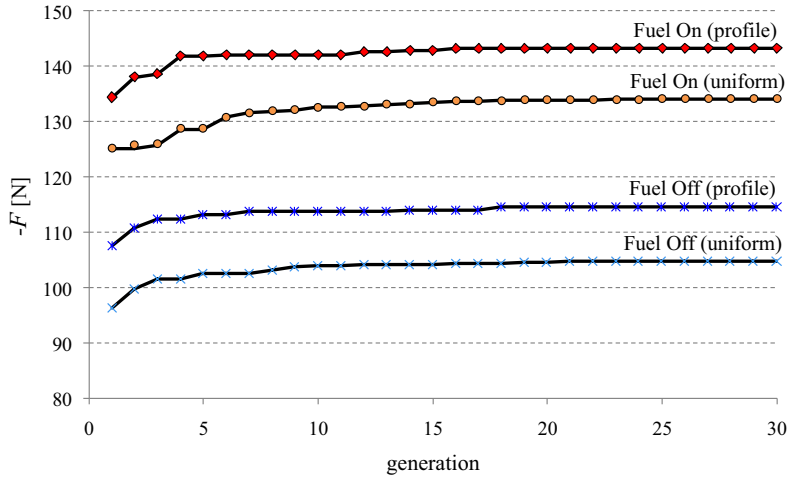


Fig. 7: Optimisation progress

Optimum geometries

The values of the objective function $(-F)$ as well as the 10 decision variables (design parameters) are plotted in Fig. 8 for the optimum geometries along with the baseline values.

Plotted and compared in Fig. 9 (a) are the optimised and baseline geometries (note that the contours are not to scale). Somewhat similar shapes with relatively straight inner and outer nozzle contours have been obtained for all four cases. A closer look in the vicinity of the nozzle entrance (Fig. 9 (b)) reveals pronounced curvature for both fuel on and off cases with inflow profiles, as also indicated by large values of the nozzle arc radius r_a in Fig. 8.

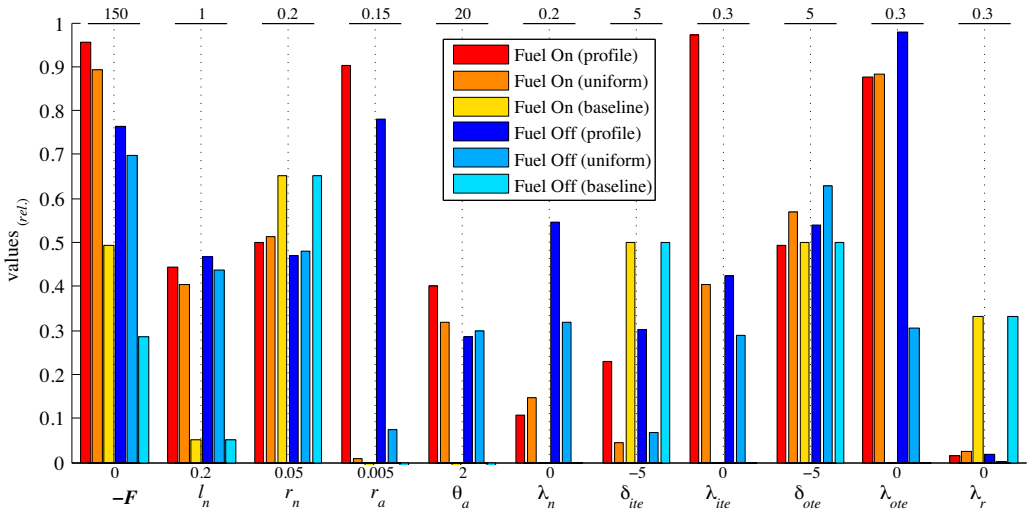
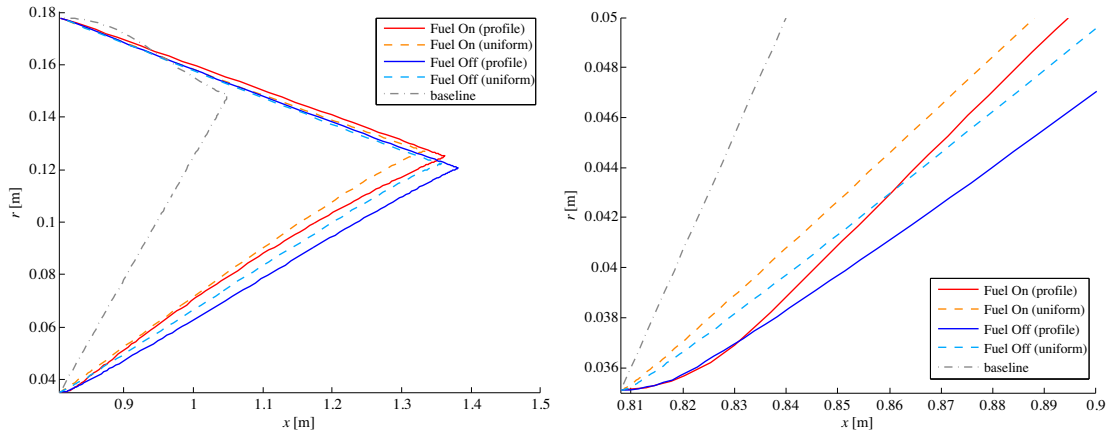


Fig. 8: Objective function and decision variable values (at 30th generation)



(a) overall

(b) close-up

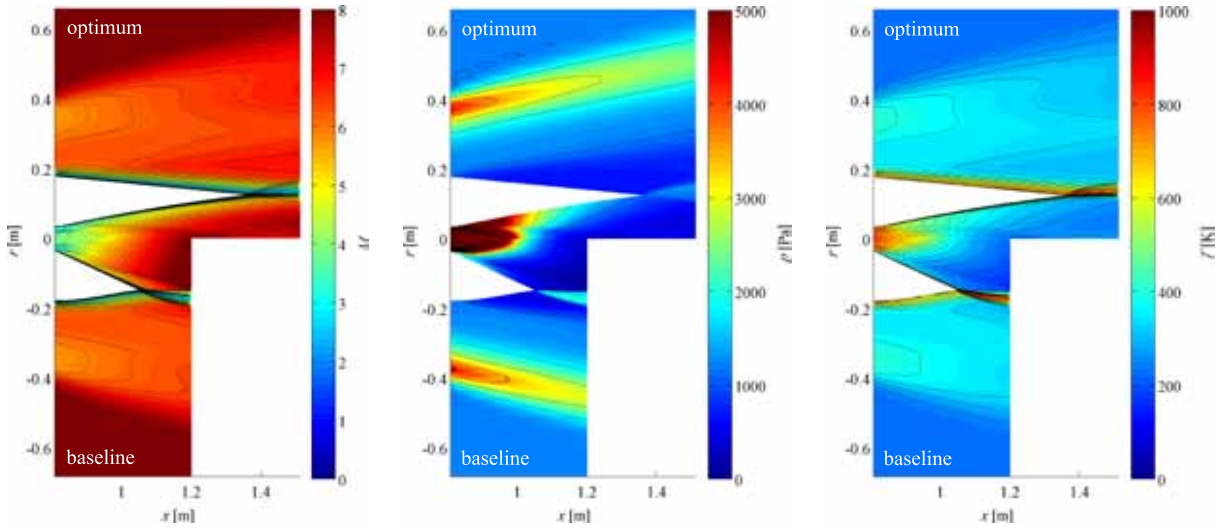
Fig. 9: Optimum and baseline nozzle geometries

Flowfields

Flow characteristics

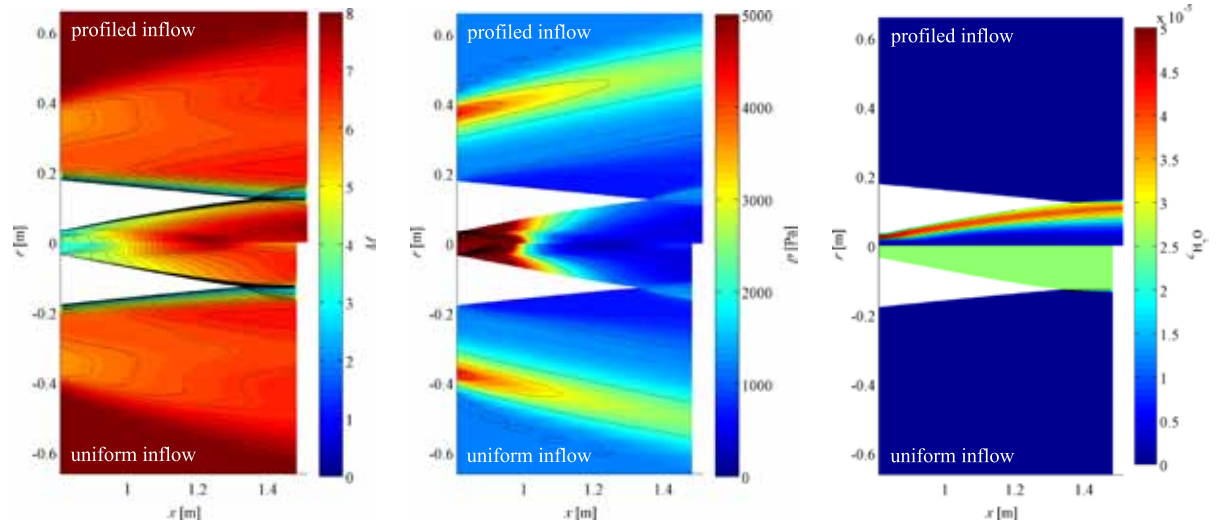
The flowfields are visualised in Fig. 10 for the nozzle geometry optimised for the profiled inflow case with fuel on, compared with the baseline flowfields. It can be noted that the optimised nozzle allows moderate expansion of the internal flow by its extended slender shape, whereas the baseline geometry experiences over-expansion due to drastic broadening of the channel, as characterised by the emergence of an incidental oblique shock to compensate for the pressure drop (at the bottom of Figs. 10 (a) and (c)). This effectively allows the near-surface high pressure to hold over a longer extent in the optimum case, which lends itself to producing greater thrust than the baseline nozzle (Fig. 10 (b)).

Compared in Fig.11 is the difference in the flowfields between the two inflow assumptions, i.e. profiled and uniform inflows, in the presence of fuel. Rather minor as the difference may



(a) Mach number (b) pressure (c) temperature

Fig. 10: Flowfields around optimum and baseline nozzles (Fuel On, profiled inflow)



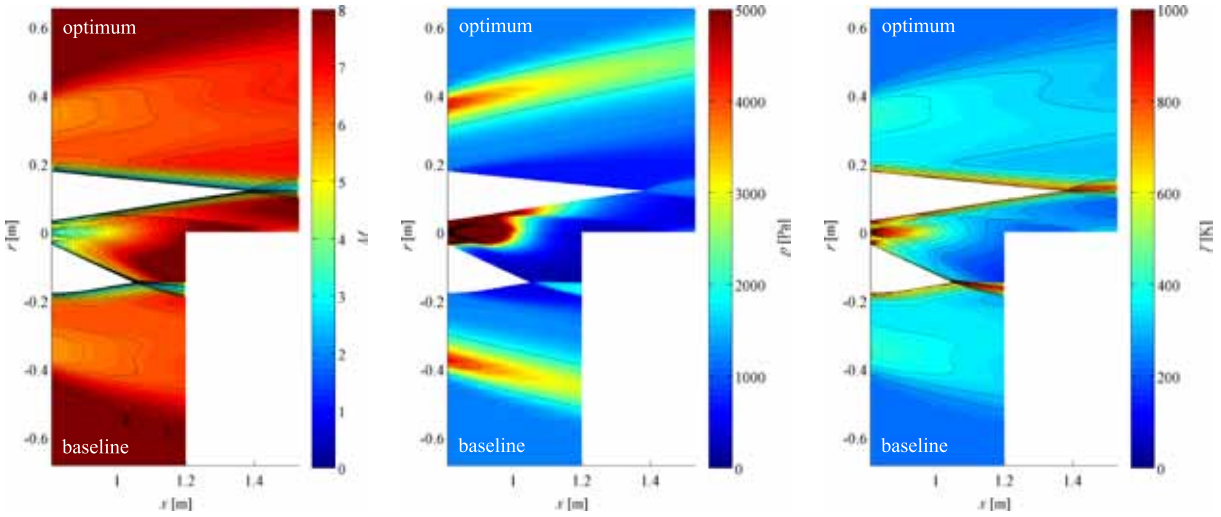
(a) Mach number (b) pressure (c) water mass fraction

Fig. 11: Flowfields around optimum nozzles with profiled / uniform inflow (Fuel On)

seem, it is noticeable in the top image of Fig.11 (a) that the incoming shock wave brought by the inflow interacts with the kernel flow in the case of the profiled inflow. In effect the high pressure that is present between the internal shock and nozzle wall is felt over a larger surface area for the profiled inflow case, as seen in Fig.11 (b)[4]. Fig.11 (c) shows the distributions of the mass fraction of the water, which is a product of hydrogen-air combustion. It can be seen that this species, which is carried over from the combustor, is smoothly diffused downstream through the nozzle.

[4]This is partly attributed to the difference in the initial expansion region as a result of optimisation, which is scrutinised later in the surface flow analysis.

The flowfields are plotted in Fig.12 for the optimum and baseline geometries with the inflow profile in the absence of fuel. It can be observed that an oblique shock wave carried from upstream (as also seen in Fig.5 (a) and (c)) reflects on the centreline and intersect with the internal shock wave (Fig.12 (a)). This interaction occurs downstream, as compared to the fuel-on case in Fig.10, causing the kernel flow to stretch farther and making the high pressure zone between the internal shock and wall more distinct (Fig.12 (b)).



(a) Mach number (b) pressure (c) temperature

Fig. 12: Flowfields around optimum and baseline nozzles (Fuel Off, profiled inflow)

Surface force

Total surface force

The force acting on an axisymmetric surface can be obtained by integrating the pressure and shear stress as follows:

$$F = \int_s (p \sin\theta - \tau \cos\theta)\, 2\pi r \, ds \tag{1}$$

The total force that acts on both sides of the nozzle, i.e. $F = F_{\text{inner}} + F_{\text{outer}}$, is compared in Fig.13 for the optimum geometries that have been achieved with three different computational mesh resolutions (as tabulated in Table 1) under the four assumed inflow conditions. Little difference is found between different mesh resolutions for all conditions, with the largest deviation of 1.1% seen for the fuel-off case with the inflow profile, which justifies the choice of the (coarse) mesh resolution adopted in this study.

Plotted in Fig.14 are the contributions of the pressure and viscous forces to the total surface force for the optimum geometries with fuel on (those for the fuel-off case are not presented here since they are similar to the fuel-on case), compared with the baseline cases. It is notable

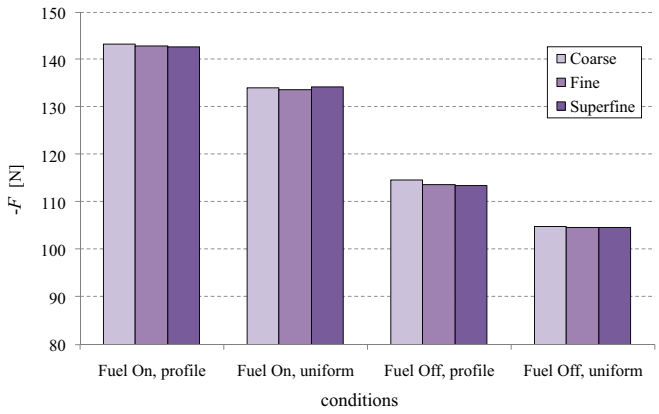


Fig. 13: Total surface force (mesh sensitivity)

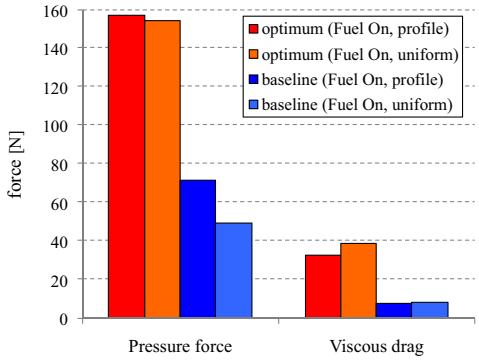
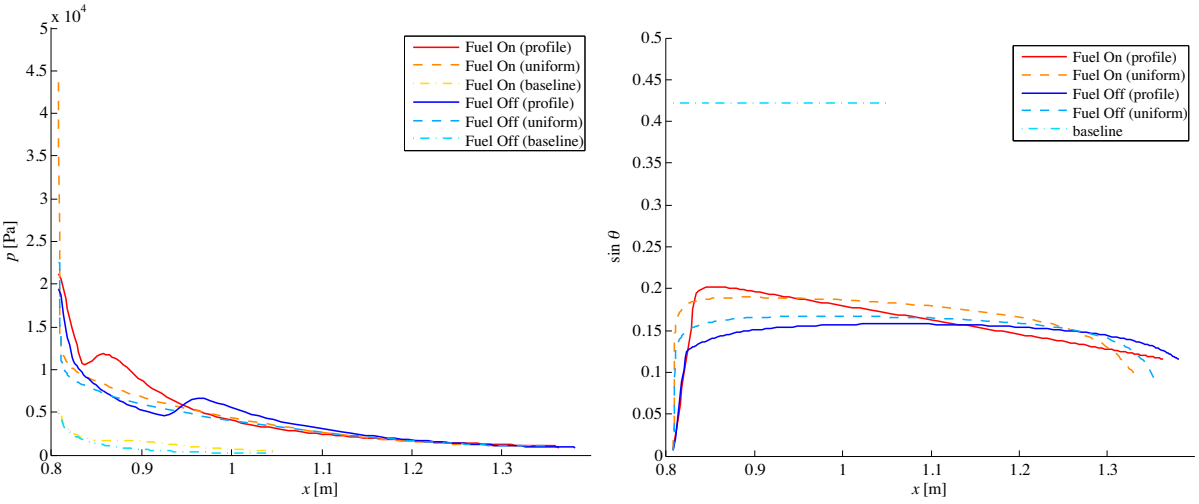


Fig. 14: Pressure / viscous breakdown of inner nozzle force (Fuel On)

that substantial increase in pressure force has been achieved by design optimisation for both cases with moderate viscous penalty, leading to appreciable gain in the overall nozzle thrust.

Surface pressure distributions



(a) surface pressure

(b) $\sin\theta$

Fig. 15: Inner nozzle surface properties

In order to investigate the driving factors of shape optimisation, the surface flowfields are examined in detail here[5]. Fig.15 shows the pressure distributions and sine variations of the surface inclination on the inner nozzle surface for the optimum and baseline geometries. High pressure level of the incoming gas emitted from the combustion chamber is maintained over extensive ranges of the inner nozzle section (Fig.15 (a)) for the optimum geometries, in contrast to the baseline nozzle geometry. The presence of a local pressure rise is indicated by a hump in the cases of profiled inflows. It is noteworthy that the initial gradients of the

[5]Only the surface pressure is discussed here because the contributions of the viscous force to the thrust have been found to be rather minor in Fig.14.

optimum contours for the profiled inflow cases are less steep than those for the uniform inflows in Fig.15 (b). This allows the surface pressure for the profiled inflow cases to vary in such a gradual manner that it increases towards the points where the localised pressure rise is present due to shock impingement. This allows the nozzle to benefit the most from high pressure in thrust production (as in $p \sin\theta$ in Eqn.(1)), without incurring losses that could arise from flow separation and associated interactions due to the adverse pressure gradients at the humps in case the gradients were too steep.

Cross reference

The performance of the nozzle geometry optimised for various inflow assumptions has been evaluated under off-design inflow conditions. The thrust force is compared in Fig. 16 for the optimum contours along with the baseline case. It can be seen that the nozzle geometries optimised for a certain inflow assumption is able to exert comparable thrust force under other flow conditions with a maximum difference of 2.5%, demonstrating the robustness of the optimised geometries as regarding thrust production.

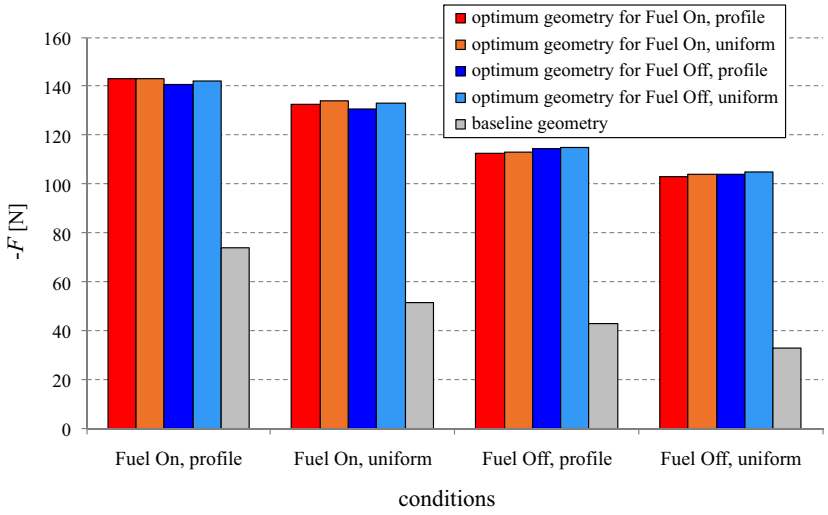


Fig. 16: Thrust force comparison in cross referencing

Conclusions

Design optimisation has been conducted to maximise the thrust of the axisymmetric scramjet nozzle for various inflow assumptions (in the presence / absence of fuel with uniform / profiled inflow) at Mach 8. The flowfields associated with the optimum geometries have been scrutinised in order to identify the key flow physics and driving factors that determine the optimium shapes and nozzle performance.

Slender nozzle contours have been yielded as a result of design optimisation under all inflow conditions considered here, maximising the benefit of high pressure nozzle inflow, compared to the baseline geometry. Profiled inflows have been found to result in curved surface in the vicinity of the nozzle entrance in order to make the most of the high surface pressure due to the impingement of the shock waves from the combustor section. The robustness of the optimum geometries in terms of thrust production has been demonstrated by cross-referencing the nozzle contours under off-design inflow conditions.

Acknowledgments

The authors are grateful to Dr. Tapabrata Ray and Amitay Isaacs at UNSW@ADFA for providing the advanced MDO capability developed in the group. We are also thankful to Mr. Yohan Alazet at École Nationale Supérieure de l'Aéronautique et de l'Espace for his assistance in developing the mesh generation and optimisation methodology.

References

[1] Paull,A., Alesi,H. and Anderson,S., "HyShot Flight Program and How it was Developed", AIAA Paper 2002-5248, Sep 2002.

[2] Boyce,R.R., Gerard,S. and Paull,A., "The HyShot Scramjet Flight Experiment – flight data and CFD calculations compared", AIAA Paper 2003-7029, Dec 2003.

[3] McClinton,C.R., "X-43 – Scramjet Power Breaks the Hypersonic Barrier: Dryden Lectureship in Research for 2006", AIAA Paper 2006-1-317, Jan 2006.

[4] Boeing, "X-51A WaveRider Breaks Record in 1st Flight", http://boeing.mediaroom.com, May 2010.

[5] Hunt, D. C., Paull, A., Boyce, R. R. and Hagenmaier, M., "Investigation of an axisymmetric scramjet configuration utilising inlet-injection and radical farming", *in Proceedings of the 19th International Symposium on Airbreathing Engines (ISABE 2009)*, Montréal, Sep 2009.

[6] McGuire, J. R., Boyce, R. R. and Mudford, N. R., "Radical farm ignition processes in two-dimensional supersonic combustion", *Journal of Propulsion and Power*, Vol. 24, No. 6, pp. 1248-1257, 2008.

[7] Östlund, J., "Flow Processes in Rocket Engine Nozzles with Focus on Flow Separation and Side-Loads", Licentiate Thesis, Royal Institute of Technology, Stockholm, 2002.

[8] CFD++, Software Package, Ver. 8.11, Metacomp Technologies, Inc., CA, 2009.

[9] Menter, F. R., "Zonal Two Equation $k-\omega$ Turbulence Models for Aerodynamic Flows", AIAA Paper 93-2906, 1993.

[10] Evans, J. S. and Schexnayder, C. J. Jr., "Influence of Chemical Kinetics and Unmixedness on Burning in Supersonic Hydrogen Flames", *AIAA Journal*, Vol. 18, No. 2, pp. 188-193, 1980.

[11] Pointwise, *Software Package*, Ver.16.02, Pointwise,Inc., TX, 2008.

[12] Mölder, S. and McGregor, R. J., "Analysis and Optimization of Scramjet Inlet Performance", ICAS Paper 90-4.7.3, 17$^{\text{th}}$ Congress of the Aeronautical Sciences, Stockholm, Sweden, Sep 1990.

[13] Ray, T. and Smith, W., "A Surrogate Assisted Parallel Multiobjective Evolutionary Algorithm for Robust Engineering Design", *Engineering Optimization*, Vol. 38, No. 8, pp. 997-1011, 2006.

[14] Ray, T., Isaacs, A. and Smith, W., "Multi-objective optimization using surrogate assisted evolutionary algorithm", (*Introduction G.P.Rangaiah*), *Multi-objective Optimization: Techniques and Applications in Chemical Engineering*, pp. 131-151, World Scientific, Singapore, 2008.

[15] Deb, K., Pratap, A., Agarwal, S. and Meyarivan, T., "A fast and elitist multiobjective genetic algorithm: NSGA-II", *IEEE Transactions on Evolutionary Computation*, Vol. 6, No. 2, pp. 182-197, 2002.

Rocket Design for Scramjet-Assisted Access-to-Space Vehicles

Thomas Ferguson
Email: thomas.ferguson@uqconnect.edu.au

Thomas Jazra
Email: t.jazra@uq.edu.au

Centre for Hypersonics
School of Mechanical & Mining Engineering
The University of Queensland
Brisbane, QLD 4072

Summary: A three-stage rocket-scramjet-rocket configuration is envisaged for incorporating scramjet propulsion into a launch vehicle intended for delivering small payloads to low Earth orbits. This paper introduces an engineering approach to the design of LH_2/LOX-fuelled first and third stages for such a vehicle, a fixed-design scramjet-powered second stage provided. The design process involves minimising the fuel mass required by stages I and III respectively while catering for the predetermined flight envelope of the airbreathing second stage. Three airbreathers, different in size and shape, are used to assemble and evaluate three rocket-scramjet-rocket configurations. The payload mass fractions determined (0.68 %, 0.76 % and 0.73 %) compare satisfactorily to those of established rocket-only launch vehicles of similar gross mass (0.4-0.7 %). The results of this study further quantify the potential of scramjets for space access, despite the current low maturity level of scramjet technology. It is shown that the configuration with the smallest airbreather achieves the highest payload mass fraction.

I. Introduction

Conventional rocket technology is currently approaching the theoretical limits of its potential. However, the cost of placing payloads into orbit remains very high and alternative launch vehicle concepts are sought after. In order to allow a significant improvement in the efficiency of access to space and lower the related cost, propulsion technologies outside of the existing rocket-based field must be pursued. One promising direction of research is the supersonic combustion ramjet (scramjet) engine [1, 2], which offers a superior specific impulse [3]. Scramjet operation requires an engine-airframe integrated vehicle design [4] which is inherently more complex than a "stand-alone" rocket propulsion system. Despite the recent success of the *X-51* waverider [5], ramjet/scramjet technology has not matured to readiness for sustained flight yet, especially for use at greater hypersonic Mach numbers. One research focus at The University of Queensland is the application of hypersonic airbreathing vehicles (HAV) in multi-stage launch configurations for the delivery of small payloads to low Earth orbits (LEO). Previous [6, 7, 8, 9] and ongoing studies feature three-stage launch vehicles comprised of rocket-powered first and third stages and a scramjet-powered second stage.

Current launch vehicles generally have more than one stage [10]. The design of the launch vehicle investigated in this work is largely driven by the second-stage HAV. Scramjet

operation is bound to a specific flight envelope, bordered by freestream dynamic pressures of approximately 20 kPa and 90 kPa [3]. Different propulsion systems are therefore required for the first and third vehicle stages. Liquid-fuelled rockets are chosen due to their technological maturity and their applicability over a large range of flight conditions.

The first stage is required to deliver the vehicle to an altitude of 26.5 km, with a flight velocity equivalent to Mach 6 and a flight path angle of 0°. At this condition, the scramjet engines of the HAV ignite to accelerate the vehicle along a constant-dynamic pressure trajectory ($q_\infty = 50$ kPa). After engine burnout the HAV conducts an unpowered pull-up manoeuvre to further gain altitude and escape from the regime of high dynamic pressure. A value of $q_\infty = 1$ kPa is targeted for deployment of the third stage, at which condition the aerodynamic forces are assumed to be negligible. The third stage then provides the final acceleration required for orbital insertion of the payload at $h = 200$ km above sea level.

The current work has two main objectives. These are

- to develop a methodology for designing the first and third rocket stages for an HAV-assisted launch vehicle, given a predetermined HAV design;
- to conservatively quantify the payload mass fractions of three particular, HAV-assisted configurations.

A brief description of the second-stage HAV is provided in Sec. II. The following two sections outline the design methodologies for the first and third stages respectively. In Sec. V the physical design of the three configurations is presented, along with a discussion of the ascent trajectories. The payload mass fractions calculated for the three launch vehicles are referenced against established rocket-only launch technology and the paper concludes with a summary in Sec. VI.

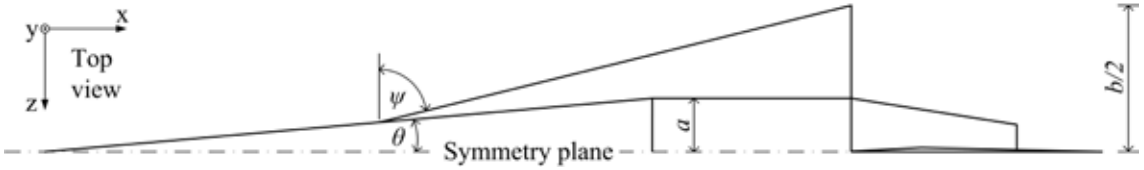
II. Second Stage: Hypersonic Airbreathing Vehicle (HAV)

The design of the HAV is derived from a winged-cone vehicle (WCV) originally developed at NASA with the intent to study single-stage-to-orbit vehicle design [11]. The vehicle was subject to a recent project[1] conducted at The University of Queensland examining different HAV designs for scramjet-assisted access to space. Each configuration integrates a modified WCV airframe with three Mach 12 REST (rectangular-to-elliptical-shape transition) scramjet engines [12]. The WCV concept was chosen as the baseline design for the second stage since it offers considerable internal volume for storage, while retaining essential characteristics of a hypersonic accelerator [13]. Three HAV configurations investigated in the aforementioned project are used as input for the current work. The design and the flight path of the three HAV remain fixed which sets the performance requirements for the first and third rocket stages. The conceptual HAV design is illustrated in Fig. 1, revealing the parameters varied independently between the three configurations, except for the inlet capture width of the scramjets. More comprehensive detail on the engine design is presented in the works of Suraweera and Smart [12], and Doherty [14]. The specific parameters of *HAV-I*, *HAV-IIa* and *HAV-IIb* are listed in Table 1.

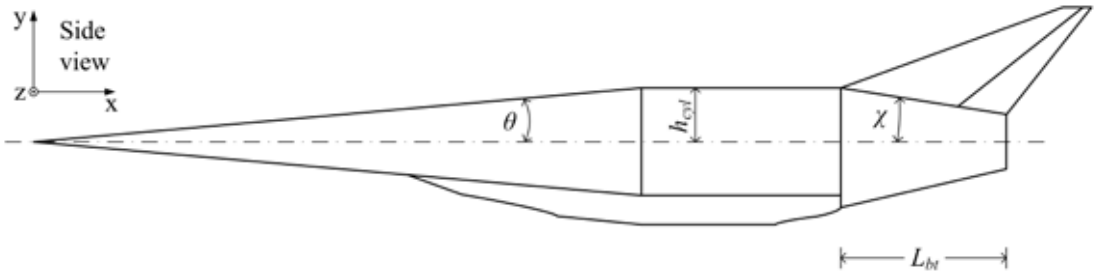
[1] "Optimisation of Hypersonic Vehicles for Airbreathing Propulsion", Ph.D. thesis submitted for examination at The University of Queensland by Thomas Jazra in November 2010.

Table 1: Design parameters of the three second-stage HAV. The value of h_{cyl} is also employed as the diameter of the first-stage rocket booster.

	h_{cyl} [m]	θ [deg]	AR [-]	χ [deg]	L_{bt} [m]	Ψ [deg]	b [m]	w_{cap} [m]	$L_{fuselage}$ [m]
HAV-I	**1.05**	6.5	1.000	19.9	2.86	78.6	4.50	0.787	22.63
HAV-IIa	**0.79**	6.4	1.065	17.0	2.31	78.2	4.08	0.595	17.26
HAV-IIb	**0.79**	5.2	1.225	10.2	2.15	81.1	3.90	0.744	20.71



(a) HAV airframe



(b) Scramjet integration

Fig. 1: HAV design parameters. The capture width of the scramjet engines, w_{cap}, is not shown; the fuselage aspect ratio, AR, is calculated as a/h_{cyl}.

III. Design of Rocket-Powered Stage I

This section discusses the methodology developed for the design of the first-stage rocket booster. This stage is required to accelerate the launch vehicle to a flight condition where the scramjet engines of the HAV can take over. This condition is referred to as *scramjet ignition* and defined by the following three parameters:

- Altitude, $h = 26.5$ km
- Mach number, $M = 6$
- Flight path angle, $\gamma = 0°$

The key objective is to reach this target frame using as little fuel as possible. This way the vehicle gross mass is kept to a minimum, which is necessary to maximise the overall launch vehicle efficiency. With the target altitude and flight Mach number set, the additional input required for the design comprises knowledge of the aerodynamic vehicle characteristics (1), the vehicle mass properties exclusive of the first stage (2), as well as an engine type to be installed on the first stage (3).

1. Aerodynamic Vehicle Properties:
The launch vehicle ascends at angles of attack between $-4°$ and $0°$, with the HAV mounted atop the first rocket stage. It is assumed that the aerodynamic drag is caused mainly by the vehicle frontal area during the ascent which implies that the total aerodynamic drag is dominated by pressure drag. Stage I acts as an extension of the fuselage of the second stage since both components have similar cross-sectional areas. The booster constitutes a significant proportion to the gross mass of the vehicle, however it barely contributes to the frontal area (given the small values of α). Following this reasoning, the influence of the aerodynamic forces and moments of the first-stage booster on the ascent trajectory is neglected; the aerodynamic properties of the launch vehicle are assumed to be those of the HAV.

Determining the aerodynamic properties of the HAV for a flight envelope spanning from takeoff to *scramjet ignition* is intricate and requires examining subsonic, transonic, supersonic and eventually hypersonic flowfields. To overcome this task, the lift and drag coefficients of the original winged-cone vehicle are adopted for the first-stage design methodology. The applicability of this simplification was assessed numerically to prevent overoptimistic conclusions or unphysical stage designs. This involved repeating the design process for stage I twice, assuming that the aerodynamic coefficients provided in Ref. [11] were either too high or low by 25 %, respectively. The discrepancies in the gross mass of the launch vehicle calculated using the modified aerodynamic coefficients are within 3 %, as presented in Table 2. This result justifies the use of the original aerodynamic dataset.

As a limitation of this study, the launch vehicle is assumed to be trimmed at all times prior to reaching the point of *scramjet ignition*. Note that not accounting for the additional lift obtained from the first stage during the ascent constitutes to a conservative estimate of the lifting capacity and hence the overall performance of the launch vehicle.

Table 2: Variation of the fuel and structural masses of stage I and the gross mass of the launch vehicle with the aerodynamic HAV properties. C_0 denotes the original dataset [11].

Aerodynamic Dataset	$m_{f,1}$ [kg]	$m_{str,1}$ [kg]	m_{gross} [kg]	Δm_{gross} [%]
C_0	8635	879	18445	
0.75 C_0	8243	839	18013	-2.3
1.25 C_0	9132	930	18993	+3.0

2. Vehicle Gross Mass:
The mass properties of the three HAV are adopted from the work of Jazra (2010). Two masses are of particular significance to the present study; i.e. the gross mass of the HAV prior to *scramjet ignition*, m_2, and the mass of the on-board fuel used during scramjet operation, $m_{f,2}$. The gross mass of the third stage, m_3 (= 1000 kg), counts towards m_2 and is also adopted and kept constant for consistency. The relevant data for the three HAV is given in Table 3, showing considerable differences in terms of the mass properties and final flight conditions. A greater mass of the HAV causes the requirement of a more powerful first rocket stage. Conversely, beneficially for the gross mass of the launch vehicle, a greater final altitude and Mach number reduce the impulse requirement of the third stage. Consequently, the three stages have to be assessed in their entirety to rate the overall efficiency of the launch vehicle.

Table 3: HAV: gross and fuel mass, altitude and Mach number at scramjet shutdown.

			Scramjet shutdown	
	m_2	$m_{f,2}$	h	M
	[kg]	[kg]	[km]	[-]
HAV-I	8381	1953	34.4	10.8
HAV-IIa	4778	640	32.4	9.3
HAV-IIb	6321	1087	33.2	9.9

3. Rocket Engine:
Five different rocket engines are selected from existing first stages based primarily on their thrust rating. Preliminary analysis of the launch vehicle indicated that a total engine thrust in the order of 500 kN would be required. To allow for sufficient variation in the flight profile rocket engines are considered with an approximate range of thrust values from 250 kN to 1000 kN. Neglecting the aerodynamic influence of the booster (and hence its length) the fuel volume is not constrained in the design process. In these circumstances liquid hydrogen is the most efficient fuel despite its low density. It has a higher energy content and a more rapid burning rate than hydrocarbons [15, 16]. The five selected engine types use combinations of liquid hydrogen (LH_2) as fuel and liquid oxygen (LOX) as oxidiser; the engine specifications are summarised in Table 4.

Table 4: Rating of rocket engines considered for installation on stage I.

Engine	$m_{eng,1}$	T	I_{sp}	m_{LOX}/m_{LH2}
	[kg]	[N]	[s]	
RL-50	500	290000	472	5.5
RL-60	499	289100	470	-
RD-57A-1	550	395000	460	5.8
Vulcain	625	1075000	431	6.2
LR87 LH2	700	587000	403	-

A. Design Methodology

A three-degree-of-freedom (3-DOF) trajectory optimisation code is developed to calculate an ascent trajectory from takeoff to *scramjet ignition* with a permitted deviation from the requested flight conditions of less than 1 %. The code optimises the flight profile for a minimal fuel mass independent of the booster's physical dimensions. The HAV dictates the takeoff load and the aerodynamic vehicle properties, and the engine selection determines the thrust, specific impulse and mass of the engine. The optimisation process involves varying three parameters: the angle of attack, α, the time at which α is adjusted, $t_{\Delta\alpha}$, and the initial fuel mass of the first stage, $m_{f,1}$ ($= m_{LH2,1} + m_{LOX,1}$). The physical dimensions of stage I are computed after the optimisation, based on $m_{f,1}$ and the following constraints: the diameter of the first stage is set to 2 h_{cyl}, and the average density is inferred from existing rocket technology. The design methodology is described next, before detail on the physical design of the stage is provided.

The ascent trajacetory of stage I is designed to reach its peak at an altitude of 26.5 km, placing the HAV at zero flight path angle. This condition can be satisfied by variation of the angle of attack during the ascent. For the scope of this project only one α-correction is permitted between vehicle takeoff and *scramjet ignition*. Launched vertically the configuration must ascend at negative angle of attack to transition into horizontal flight as it approaches the target frame. "Increasing" (towards higher negative values) the angle of attack early results in higher aerodynamic drag in the dense part of the atmosphere thus lowering the final vehicle velocity for a given fuel mass. Conversely, adjusting the angle of attack late necessitates a comparatively steep α-value to prevent the vehicle from overshooting and missing the targeted flight condition. The steep angle of attack required in the case of a late correction produces a significant increase in drag, as does an early correction, indicating the presence of an optimum solution between these extremes. Including $m_{f,1}$ in the optimisation ensures that sufficient fuel is carried on board to reach the target frame while it holds the design from carrying excess fuel upon reaching the trajectory peak. Engine burnout of the first stage is assumed to occur before the vehicle reaches this peak allowing for a design coasting phase of approximately 10-20 seconds.

The trajectory is calculated by integrating the equations of planar motion using a fourth-order Runge-Kutta scheme to update the aerodynamic and propulsive vehicle properties at each time step (0.1 seconds). The optimisation is carried out utilising a bisection method. The industry-standard trajectory software Computer Aided Design of Aerospace Concepts (*CADAC*) [17] is used to verify the accuracy of the computed results. Fig. 2 compares the flight path angle for the ascent of stage I (shown for the *HAV-I*-based launch vehicle) computed using both the developed code and *CADAC* with discrepancies of less than 1 %. The observed accuracy is representative of all cases studied.

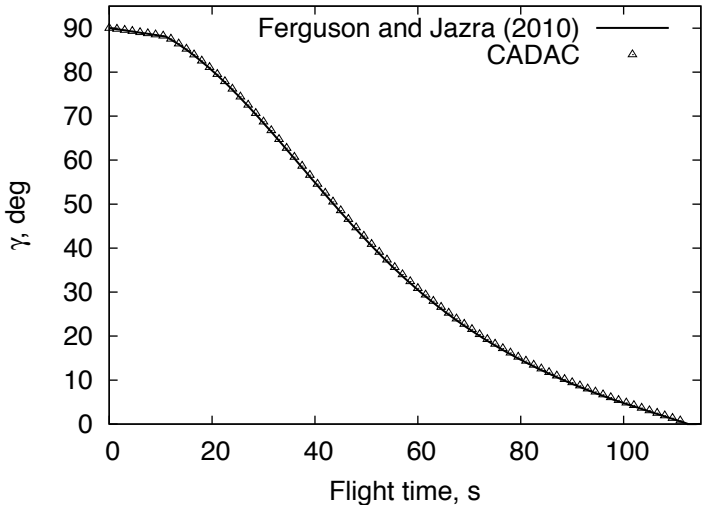


Fig. 2: Comparison of flight path angles during the ascent of stage I between the developed code and CADAC [17].

With the total propellant mass known the physical design of the first-stage booster can be developed. Keeping the diameter of the booster fixed, the parameters to be determined include: oxidiser/fuel mass ratio, $m_{LOX,1}/m_{LH2,1}$; fuel volume, $V_{LH2,1}$, and oxidiser volume, $V_{LOX,1}$; structural mass, $m_{str,1}$; average density, ρ_1; and length, L_1.

Table 5: Specifications of existing first-stage rockets [18].

Vehicle	Engine	$m_{eng,1}$ (n) [kg]	m_1 [kg]	$m_{1,\text{dry}}$ [kg]	V_1 [m^3]
H-2-1	LE-7	1714 (1)	98100	11900	351.9
H-2A-1	LE-7A	1800 (1)	113600	13600	467.5
H-2A LRB	LE-7A	1800 (2)	117000	17800	461.2
Ariane H155	Vulcain	625 (1)	170800	12700	714.1
Ariane 5 HPC	Vulcain 2	811 (1)	186000	12700	714.1
Energia M	RD-0120M	3450 (1)	272000	28000	931.3
Saturn II	J-2	1438 (5)	490778	39048	1974.4
Saturn II-SL	J-2-SL	1360 (5)	490952	44240	1974.4
Saturn II-INT-17	HG-3-SL	1400 (7)	495000	48000	1974.4
Titan 5	PW 1M	4500 (1)	500000	45000	1696.5
Ares	SSME	3177 (3)	787700	64200	2556.2
Energia Core	RD-0120	3450 (4)	905000	85000	2772.4
Cargo LV	SSME	3177 (5)	1093330	88449	3561.7

The oxidiser/fuel mass ratio is adopted from the engine data provided in Table 4. Based on this ratio and the densities of $\rho_{LH2} = 71$ kg/m^3 and $\rho_{LOX} = 1141$ kg/m^3 the volumes of the propellant components are calculated. The structural mass is estimated by deriving a relation between the fuel and structural masses of existing first-stage rockets in the form: $m_{f,1}/m_{str,1} = \text{f}(m_1)$. Table 5 shows a breakup of the total mass for a selection of LH_2/LOX-fuelled booster stages. The fuel/structure mass ratio is determined as:

$$\frac{m_{f,1}}{m_{str,1}} = \frac{m_1 - m_{1,\text{dry}}}{m_{1,\text{dry}} - n \times m_{eng,1}} \tag{1}$$

where n is the number of engines installed on the stage respectively. To establish a relation between $m_{f,1}/m_{str,1}$ and m_1, a linear fit through the ratios devised from Table 5 is computed, as shown in Fig. 3(a). A linear fit is chosen specifically as it reduces the impact of the three outliers having ratios greater than 13 at a total mass below 500 tons. Assuming a gross mass of the first stage between 6000 kg and 10000 kg yields a value of $m_{f,1}/m_{str,1} \approx 9.82$-$9.84$.

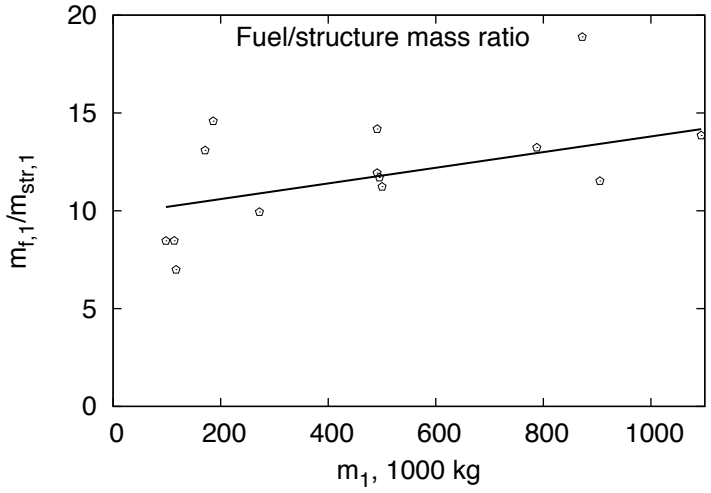
The average density of stage I is determined in a similar way as the fuel/structure mass ratio, as shown in Fig. 3(b). The mass of each engine is fixed and does not scale with the volume of the stage, it is not included in the average density, so that:

$$\rho_1 = \frac{m_1 - n \times m_{eng,1}}{V_1} \tag{2}$$

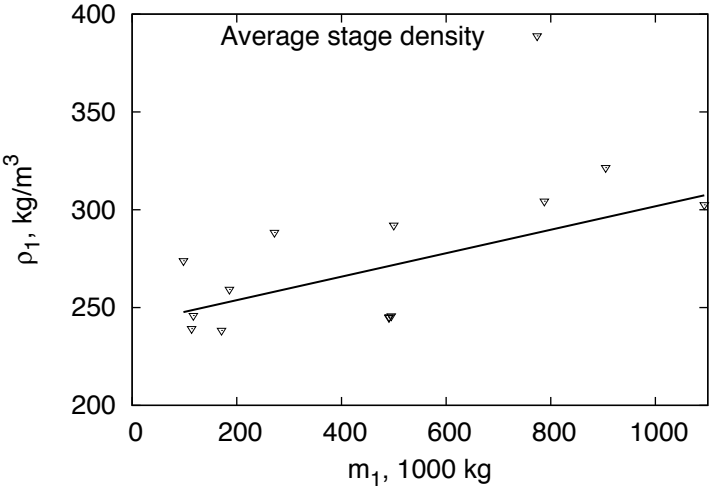
where V_1 is the volume of the cylinder bounding the stage, not accounting for wall thicknesses. Based on an estimated gross mass of 6000-10000 kg a density of stage I of $\rho_1 \approx 242$ kg/m^3 is calculated. The length of the stage is calculated from ρ_1, the diameter of the stage and the length of the engine nozzle.

IV. Design of Rocket-Powered Stage III

This section introduces the third vehicle stage. At takeoff, stage III is accommodated inside the fuselage of the HAV. After deployment this stage provides the final impulse for insertion



(a) Mass ratio



(b) Density

Fig. 3: Fuel/structure mass ratio and average density of existing first-stage rocket boosters using LH_2/LOX (data taken from Table 5).

of the payload into a circular orbit ($h = 200$ km). The key objective for the design is to maximise the orbital payload mass while minimising the fuel mass required for the final ascent. Detail on the developed methodology, specified constraints and the physical design of the third stage is presented below.

A. Design Methodology

The design of stage III is constrained by the internal volume of the fuselage of the HAV. For the scope of this work a bounding cylinder with a fixed geometry[2] of 4.50 m $\times$ 1.10 m (length $\times$ diameter) is allocated for stage III. The total mass of the stage is constrained to

[2]The constraints on the dimensions and the mass of the stage III are adopted from the work of Jazra (2010).

1000 kg, including fuel, oxidiser, engine, structure and payload. The design methodology incorporates the following three steps: unpowered ascent of the HAV to escape from the sensible atmosphere (1); deployment of the third vehicle stage (2); rocket-powered final ascent to reach the target orbit (3).

1. Unpowered Ascent
Staging manoeuvres conducted within the effective atmosphere often experience critical aerodynamic loads [19, 20]. The design of stage III builds on the final flight conditions of the HAV that terminates its powered flight at $q_\infty \approx 50$ kPa. To leave the high-q_∞ regime the vehicle performs an unpowered pull-up manoeuvre after engine burnout aimed to lower the freestream dynamic pressure to a non-critical level. It is assumed that the aerodynamic effects are of negligible order at this flight condition.

The unpowered ascent of the HAV is modelled using *CADAC* while the aerodynamic vehicle properties are computed using *HYPAERO*. During this unpowered period, the vehicle follows a parabolic trajectory. Initiated by an increase in the angle of attack the craft decelerates as it gains altitude until reaching a peak, the *deployment point*, and accelerates again and loses altitude under the influence of gravity upon passing this peak. The value of α is held constant for the duration of the unpowered ascent. At the *deployment point* the HAV reaches its maximum altitude and minimum flight velocity, so the freestream dynamic pressure is minimal. The optimum angle of attack is that for which the HAV incurs the smallest loss in kinetic energy to reach a peak where $q_\infty = 1$ kPa.

Fig. 4: Open hatch prior to the deployment of stage III.

2. Deployment of Stage III
The mechanism for releasing the third vehicle stage has been investigated previously [21] with a summary given here. Prior to deployment, stage III is housed in a frame on sliding rails with the frame being attached to the underside of a rear-opening hatch. On deployment the hatch is opened (see Fig. 4) by a critically-damped spring system, utilising helical compression springs. The release is initiated by pyro-assisted clamps on the inner surface of the hull and the hatch is retracted by hydraulic actuators after the deployment. The mechanism ejects the third stage rearwards, placing it to a nominal clearance of 15 m within 10 seconds. The resulting velocity change of the HAV is considered to be negligible with respect to the inertial velocity of the craft and not accounted for. More comprehensive detail on the deployment mechanism is provided in Ref. [21].

3. Rocket-powered Ascent to LEO

Stage III must attain sufficient velocity to allow for a stable orbit at $h_{\text{orbit}} = 200$ km above sea level. The flight path for the final climb is calculated independent of the physical size of the stage, which is reasonable in the close-to-zero-atmosphere environment beyond the *deployment point*. Three unknowns are solved for: the ideal velocity increase required for the transfer from $h_{\text{deployment}}$ to h_{orbit}, Δv_{ideal}; the loss in velocity due to gravity incurred during the ascent, Δv_{loss}; and the fuel mass required, $m_{f,3}$ $(= m_{LH2,3} + m_{LOX,3})$.

The ideal velocity increase is calculated as the difference between the inertial velocity of the stage upon deployment and the final velocity required by the payload to sustain a circular target orbit. The solution is based on a Hohmann transfer which assumes two ideal impulsive burns to shift the vehicle from one circular orbit to another. The two velocity gains, firstly to transition from the *deployment point* into an elliptical transfer orbit with an apogee of $R_E + 200$ km, and secondly to circularise the orbit, are determined using the *Vis-viva* equation [22]:

$$\Delta v_{\text{ideal}}^2 = GM \left(\frac{2}{r} - \frac{1}{a} \right) \tag{3}$$

where G is the universal gravitational constant, M the mass of the Earth, r the distance from the orbit's focus, and a the semi-major axis (for circular orbits: $r = a$). However, the large initial increase from approximately 3 km/s to about 8 km/s (at the *deployment point*, the HAV is not in a circular orbit) cannot be achieved by an ideal impulsive burn. A longer engine burn is therefore modelled that accounts for the gravitational losses during the acceleration period using the following equation adapted from Ref. [3]:

$$\frac{m_{f,3}}{m_3} = 1 - \exp\left[-\frac{\Delta v}{I_{sp} g_0} - \frac{1}{I_{sp} g_0} \int_0^t \left(\frac{D}{m_3} + g \sin\gamma \right) \mathrm{d}t \right] \tag{4}$$

Assuming the marginal aerodynamic drag to be zero, Eq. 4 yields the fuel mass required for the extended burning period.

The following parameters can then be determined to finalise the design: engine thrust, T_3, and specific impulse, $I_{sp,3}$; oxidiser/fuel mass ratio, $m_{LOX,3}/m_{LH2,3}$; fuel volume, $V_{LH2,3}$, and oxidiser volume, $V_{LOX,3}$; structural mass, $m_{str,3}$; average density, ρ_3; and length, L_3.

The vacuum-rated thrust and specific impulse of the engine are devised by scaling from previous data for upper-stage rockets (see Table 6). A relatively high mass proportion of 10 % of the allocated total mass of stage III is chosen for the engine, $m_{eng,3} = 100$ kg, in order to provide for the substantial impulse requirement between the *deployment point* and the target orbit. Optimisation of this proportion in the future may allow for enhanced efficiency of the stage. A scaled-down version of Pratt & Whitney's RL-10A-3A is used, retaining the original oxidiser/fuel mass ratio of 5 [18] while reducing the original mass of 141 kg to the design value of 100 kg. An exponential fit is applied to determine the value of T_3, and a linear fit is used to determine $I_{sp,3}$.

The structural mass and the average density of stage III are calculated in the same way as the according properties of stage I but based on the data in Table 6. Computing linear fits, the approximate mass of stage III of 800 kg (excluding engine and payload) produces a fuel/structure mass ratio of 6.5 and an average density of 240 kg/m^3. The overall length of the stage is calculated from the average density, the preset stage diameter and the length of the engine nozzle.

Table 6: Specifications of existing upper-stage rockets [18].

Vehicle	Engine	$m_{eng,3}$ (n) [kg]	m_1 [kg]	$m_{3,\text{dry}}$ [kg]	V_1 [m^3]	T [N]	I_{sp} [s]
Ariane H8	HM7-A	149 (1)	9687	1457	42.4	67100	443
Ariane H10	HM7-B	155 (1)	12000	1600	49.1	70000	447
Ariane H10+	HM7-B	155 (1)	12800	1740	50.6		
Ariane H10-3	HM7-B	155 (1)	12310	1570	50.9		
Ariane 5 ESC A	HM7-B	155 (1)	16500	2100	125.0		
Centaur C	RL-10A-1	131 (2)	15600	1996	41.2	66700	425
Centaur D/E	RL-10A-3	131 (2)	16258	2631	44.5	65600	444
Centaur I	RL-10A-3A	141 (2)	15600	1700	46.4	73400	444
Centaur G Prime	RL-10A-3A	141 (2)	19501	3000	83.2		
Centaur G STS	RL-10A-3A	141 (2)	16327	2600	40.2		
Centaur II	RL-10A-3A	141 (2)	18833	2053	53.4		
Centaur IIA	RL-10A-4	168 (2)	19073	2293	49.6	92500	449
Centaur G	RL-10A-3A	141 (2)	23880	2775	85.1		
N1 Block R	RD-56	282 (1)	23000	4300	68.6	69600	462
LE-5	LE-5	245 (1)	10600	1800	33.2	103000	450
LE-5EC	LE-5A	242 (1)	16700	2700	82.9	121000	452
LE-5B	LE-5B	269 (1)	19600	3000	63.9	137000	447
Ariane 5 ESC A	Vinci	280 (1)	27500	3400	93.2	180000	467

V. Results & Discussion

LH_2/LOX-fuelled first and third rocket stages were designed for each of the three HAV configurations to complement the targeted rocket-scramjet-rocket architecture. The example of the *HAV-I*-based vehicle is used in this section for a discussion of the computational results. The results for all three launch vehicles are contained in Table 8 and subdivided into results for stage I, the unpowered ascent of the HAV, stage III, and the payload mass fraction of each three-stage configuration.

A. Stage I

Five existing rocket motors were considered for application in the first-stage booster with a focus on two key characteristics; the engine-specific impulse, and the engine thrust. Rocket engines with a high specific impulse require a lower initial fuel mass and vice versa. However, the specific impulse typically varies considerably between sea level and vacuum conditions, therefore the particular application range of a rocket engine plays an important role for the selection of the engine. Engines with higher thrust levels accelerate a given mass in a shorter time period reducing gravitational and aerodynamic losses and therefore require less fuel.

The low-T engines considered in this study, RL-50 and RL-60 (Table 4), do not provide sufficient thrust to deliver the *HAV-I* and *HAV-IIb*-based launch vehicles to the point of *scramjet ignition*; the mid-sized engine, RD-57A-1, is powerful enough to accomplish the task. The two larger engines have a lower I_{sp} than RD-57A-1, which results in a significant

Table 7: Launch vehicle gross masses based on first-stage engine selection.

Engine	*HAV-I*-based [kg]	*HAV-IIa*-based [kg]	*HAV-IIb*-based [kg]
RL-50	-	10730	-
RL-60	-	10750	-
RD-57A-1	18845	10743	13748
LR87 LH2	19986	11977	15145
Vulcain	18467	11141	14089

drop in engine efficiency. This loss is not compensated by the higher thrust level of the Vulcain and LR87 LH2 engines and as such these engines require a significantly higher fuel mass in most cases. Consequently, the RD-57A-1 rocket engine is selected for use in all three launch vehicle designs providing comparatively good performance and allowing for consistency between the different results. Note that the relative mass of this engine is higher than in common rocket-only configurations [10], which leaves some room for optimisation and further constitutes to a conservative estimate of the vehicle performance. Table 7 lists the gross masses calculated for each of the three launch vehicles in combination with each of the five rocket engines considered.

In line with the two other cases investigated the fuel mass of the first stage of the *HAV-I*-based launch vehicle is of similar magnitude as the gross mass of the HAV ($m_{f,1}/m_2 \approx 1.03$). The entire fuel is consumed during a burning time of the rocket engine of 99 s. The adjustment of the angle of attack from 0° to -1.0° approximately 11 s after takeoff leaves the vehicle with a coasting period of 13 s, before the flight conditions of *scramjet ignition* are reached. With a length of 11.3 m, the stage is about half as long as the HAV.

B. Unpowered Ascent of the HAV

The unpowered ascent of the HAV after scramjet shutdown is required to place the vehicle in a close-to-zero atmosphere environment to allow the deployment of stage III. Optimising this manoeuvre involved determining the optimal angle of attack for the climb to minimise the loss in kinetic energy of the vehicle. The required angle of attack is largely determined by the kinetic energy of the craft upon scramjet shutdown and the aerodynamic efficiency (lift-to-drag ratio, L/D) of the HAV. It is in this portion of the flight path that high L/D markedly improves the payload mass fraction of the launch vehicle. In the case of the *HAV-I*-based configuration, a value of 16.2° was calculated to cause the minimum energy loss for the unpowered ascent to a ($q_\infty = 1$ kPa)-flight condition. During this phase of the trajectory *HAV-I* climbs from an altitude of 34.4 km to 56.9 km while decelerating from approximately 3.8 km/s (inertial velocity) to 3.2 km/s.

C. Stage III

The flight sequence of stage III consists of two thrust periods; an initial long burning period to accelerate into an orbital transfer trajectory, followed by an impulsive burn at the

*Table 8: Design and flight path parameters of the HAV-I, HAV-IIa and HAV-IIb-based launch vehicles. *A scaled-down version of the original RL-10A-3A engine is used.*

Stage I	Engine	$m_{str,1}$ [kg]	$m_{eng,1}$ [kg]	$m_{f,1}$ [kg]	α [deg]	$t_{\Delta\alpha}$ [s]	t_{burn} [s]	t_{coast} [s]	L_1 [m]
HAV-I	RD-57A-1	879	550	8635	-1.0	11	99	13	11.3
HAV-IIa	RD-57A-1	501	550	4915	-3.2	17	75	19	11.5
HAV-IIb	RD-57A-1	636	550	6241	-2.3	13	90	18	14.6
Stage II	*Scramjet shutdown*			*Glide*		*Deployment point*			
	$m_{dry,2}$ [kg]	h [km]	M [-]	α [deg]	t_{glide} [s]	h [km]	M [-]	v_{inert} [m/s]	
HAV-I	6432	34.4	10.8	16.2	95	56.9	9.4	3227	
HAV-IIa	4135	32.4	9.3	18.3	92	54.3	7.6	2716	
HAV-IIb	5233	33.2	9.9	18.4	93	55.4	7.6	2930	
Stage III	Engine	$m_{str,3}$ [kg]	$m_{eng,3}$ [kg]	$m_{f,3}$ [kg]	m_{pl} [kg]	L_3 [m]	$\Delta v_{3,1}$ [m/s]	$\Delta v_{3,2}$ [m/s]	
HAV-I	RL-10A-3A*	103	100	672	***125***	4.0	4687	43	
HAV-IIa	RL-10A-3A*	109	100	709	***82***	4.2	5199	44	
HAV-IIb	RL-10A-3A*	107	100	693	***100***	4.1	4984	43	
Launch vehicle								m_{gross} [kg]	m_{pl} [%]
HAV-I								18445	***0.68***
HAV-IIa								10743	***0.76***
HAV-IIb								13748	***0.73***

apogee to stabilise into a circular target orbit. The fuel mass required for this procedure is dependent primarily on the kinetic energy of the craft rather than the altitude where stage III is ignited. The magnitude of the perigee boost required to establish the transfer orbit is based on the kinetic energy of the HAV at the *deployment point*. The burning time determined for the initial boost of stage III of the *HAV-I*-based configuration is approximately 45 s, which emphasises the need to account for gravitational losses during the acceleration period (rather than assume an impulsive burn). The apogee impulse requires a burning time of only about 0.2 s, leaving the vehicle at a speed of 7.78 km/s that is required to sustain a stable orbit at 200 km above sea level. All three designs were constrained to reach this final state.

The complete altitude and velocity profiles for the ascent of the *HAV-I*-based three-stage configuration from the launch pad to the target orbit are shown in Fig. 5. After a burning time of 99 s of stage I the configuration coasts for about 13 s to reach the point of *scramjet ignition*. In progress the HAV accelerates along a constant-q_∞ flight path up to Mach 10.8 and an altitude of 34.4 km; this scramjet-powered flight segment comprises a period of 569 s. Following the unpowered ascent of the HAV to the *deployment point* at an altitude of 56.9 km and the subsequent release of stage III a stable target orbit is established approximately 2700 s after takeoff.

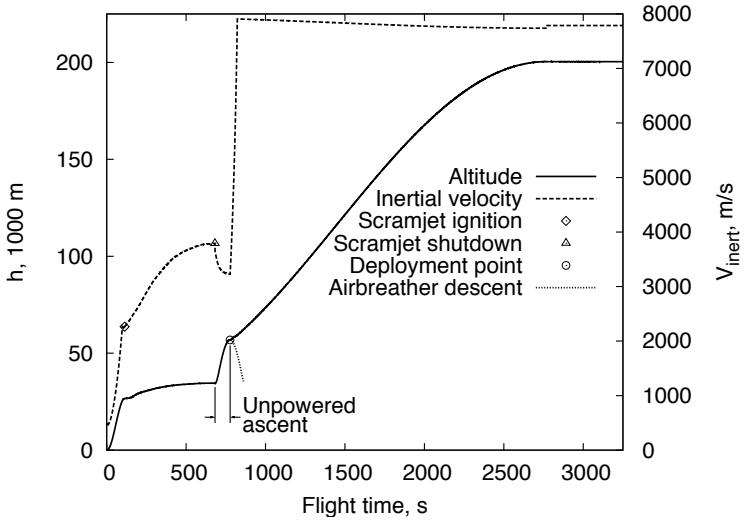


Fig. 5: Ascent of the HAV-I-based launch vehicle.

D. Launch Vehicle Efficiency

The efficiency of a launch vehicle can be expressed as the ratio of the payload mass delivered to a particular orbit to the gross mass of the vehicle. The payload mass fractions presented next are considered to be conservative performance estimates due to the following assumptions incorporated into the analysis:

- The lift generated by the cylindrical shape of the first stage during the ascent to the point of *scramjet ignition* was not accounted for.
- The terminal flight path angle for the first-stage ascent was commanded to be zero degrees, however the HAV performed an immediate pull-up upon *scramjet ignition* to accelerate at positive angle of attack. Thus, two drag penalties were incurred which could be avoided by streamlining the two attitude manoeuvres.
- Comparatively high engine-mass proportions were assumed for stages I and III to ensure the flight conditions dictated by the HAV were catered for. Optimisation of the two masses may allow for lower engine-mass proportions and hence higher relative fuel and/or payload masses.
- The mass of the third stage, m_3, was adopted from the work of Jazra (2010). Possible improvement of the launch vehicle efficiency through variation of m_3 should be investigated in the future; however, this will require recalculating the performance of the HAV accordingly.

The payload mass fractions are plotted as a function of the gross mass for the three launch vehicles investigated in this project in Fig. 6 and compared to those of several rocket-only configurations [10]. The airbreather-assisted launch vehicles have gross masses of approximately 18.4 (*HAV-I*), 10.7 (*HAV-IIa*) and 13.7 (*HAV-IIb*) tons which places them at the bottom end of the scale shown. A linear trendline is included for the rocket-only configurations; however the trend approximately agrees with the predicted vehicle efficiencies of the airbreather-assisted vehicles of 0.68 % (*HAV-I*), 0.76 % (*HAV-IIa*) and 0.73 % (*HAV-IIb*) as well. While the payload mass fraction of rocket-only launch vehicles typically increases with gross mass [10] an accordingly conclusive prediction for the rocket-scramjet-rocket

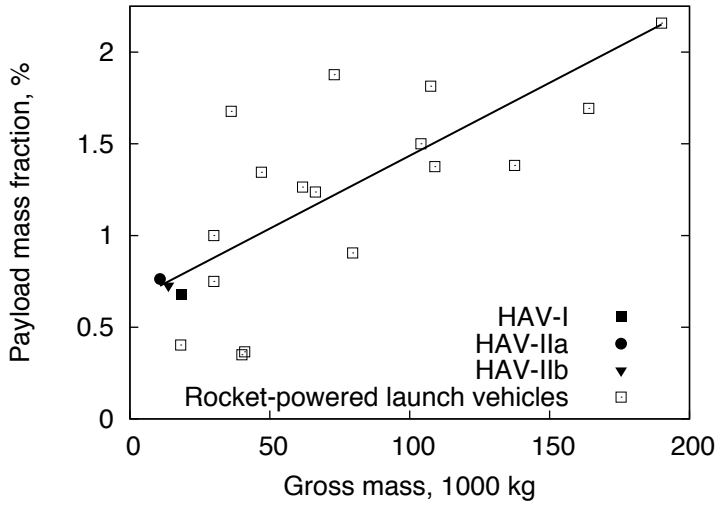


Fig. 6: Comparison between scramjet-assisted and rocket-only launch vehicles.

configurations is not possible at this stage. The three designs investigated are too similar in terms of their takeoff mass, and also a greater number of sample vehicles would be desirable. Nonetheless, the proposed rocket-scramjet-rocket architecture appears to be a competitive alternative for the future design of launch vehicles for orbiting small payloads, especially considering the assumptions in the performance analysis stated above. Moreover the fact that the lightest design achieves the relatively best overall performance is a trend worth investigating further.

VI. Summary

An engineering approach to the sizing of rocket-powered first and third stages was presented, aimed to complement the design of rocket-scramjet-rocket launch vehicles for the delivery of small payloads to LEO. The design and the flight envelope of the airbreathing second stage were fixed. An unpowered flight period of the burnt-out second stage was included in the ascent trajectory prior to the release of stage III to reduce the freestream dynamic pressure to a non-critical level. Due to this manoeuvre in particular an aerodynamic shape of the second stage aids to contribute to a high efficiency of the launch vehicle.

The developed methodologies for the design of stages I and III model a rocket as an assembly of fuel and oxidiser tanks, external structure, and engine including thrust nozzle; an additional payload mass is accounted for in the design of stage III. The stages are optimised for minimum fuel consumption thereby maximising the payload mass fraction for the mission profile specified.

The design methodologies provide sufficient flexibility for the investigation of launch vehicles of different scale. However in this study three scramjet-assisted vehicles of similar scale were investigated. The payload mass fractions of these configurations (0.68 %, 0.76 % and 0.73 %) compare satisfactorily to those of existing rocket-only configurations of similar gross mass (0.4-0.7 %). More advanced modelling in particular of the minimum required engine masses may allow for further improvement of these vehicle efficiencies.

A greater number of sample vehicles with a larger range of gross masses will be desirable in the future to study scaling trends of rocket-scramjet-rocket launch vehicles.

References

[1] Curran, E. T., Heiser, W. H., and Pratt, D. T., "Fluid Phenomena in Scramjet Combustion Systems," *Annual Review of Fluid Mechanics*, Vol. 28, 1996, pp. 323–360.

[2] Billig, F. S., "Research on Supersonic Combustion," *Journal of Propulsion and Power*, Vol. 9, No. 4, 1993, pp. 499–514.

[3] Heiser, W. H. and Pratt, D. T., *Hypersonic Airbreathing Propulsion*, American Institute of Aeronautics and Astronautics, Washington, 1994.

[4] Javaid, K. H. and Serghides, V. C., "Airframe-Propulsion Integration Methodology for Waverider-Derived Hypersonic Cruise Aircraft Design Concepts," *Journal of Spacecraft and Rockets*, Vol. 42, No. 4, 2005, pp. 663–671.

[5] ScienceDaily, "http://www.sciencedaily.com/releases/2010/05/100527130438.htm," 2010.

[6] Smart, M. K. and Tetlow, M. R., "Orbital Delivery of Small Payloads Using Hypersonic Airbreathing Propulsion," *Journal of Spacecraft and Rockets*, Vol. 46, No. 1, 2009, pp. 117–125.

[7] Jazra, T. and Smart, M. K., "Comparison of Two Optimisation Approaches for the Design of Hypersonic Airbreathing Accelerators," *Proceedings of the 9th Australian Space Science Conference*, Sydney, Australia, 2009, pp. 163–174.

[8] Jazra, T. and Smart, M. K., "Preliminary Design of an Integrated Scramjet-Powered Accelerator," AIAA Paper 2009-7418, 2009.

[9] Razzaqi, S., Jazra, T., Smart, M. K., and Ferguson, T., "Optimisation of Scramjet-Assisted Access-to-Space Vehicles Using Oxygen Enrichment," *10th Australian Space Science Conference*, Brisbane, Australia, 2010.

[10] Isakowitz, S. J., Hopkins, J. B., and Hopkins, J. P., *International Reference Guide to Space Launch Systems*, American Institute of Aeronautics and Astronautics, Reston, VA, 2004.

[11] Shaughnessy, J. D., Pinckey, S. Z., McMinn, J. D., Cruz, C. I., and Kelley, M.-L., "Hypersonic Vehicle Simulation Model: Winged-Cone Configuration," NASA TM 102610, 1990.

[12] Suraweera, M. V. and Smart, M. K., "Shock-Tunnel Experiments with a Mach 12 Rectangular-to-Elliptical Shape-Transition Scramjet at Off-Design Conditions," *Journal of Propulsion and Power*, Vol. 25, No. 3, 2009, pp. 555–564.

[13] Curran, E. T. and Murthy, S. N. B., *Scramjet Propulsion*, American Institute of Aeronautics and Astronautics, Reston, VA, 2000.

[14] Doherty, L., *Investigation of Thrust Generation in 3D Scramjets for Access-to-Space Applications*, Confirmation of Ph.D.-Candidature Report, The University of Queensland, Brisbane, Australia, 2009.

[15] Mehta, U. B. and Bowles, J. V., "Two-Stage-to-Orbit Spaceplane Concept with Growth Potential," *Journal of Propulsion and Power*, Vol. 17, No. 6, 2001, pp. 1149–1161.

[16] Kirkham, F. S. and Hunt, J. L., "Hypersonic Transport Technology," *Acta Astronautica*, Vol. 4, No. 1-2, 1977, pp. 181–199.

[17] Zipfel, P. H., *Modeling and Simulation of Aerospace Vehicle Dynamics*, American Institute of Aeronautics and Astronautics, Reston, VA, 2007.

[18] Wade, M., "Encyclopedia Astronautica (http://www.astronautics.com/)," 2009.

[19] Moelyadi, M. A., Breitsamer, C., and Laschka, B., "Stage-Separation Aerodynamics of Two-Stage Space Transport Systems Part 1: Steady-State Simulation," *Journal of Spacecraft and Rockets*, Vol. 45, No. 6, 2008, pp. 1230–1239.

[20] Moelyadi, M. A., Breitsamer, C., and Laschka, B., "Stage-Separation Aerodynamics of Two-Stage Space Transport Systems Part 2: Unsteady Simulation," *Journal of Spacecraft and Rockets*, Vol. 45, No. 6, 2008, pp. 1240–1250.

[21] Ferguson, T., *Small Satellite Separation from a Hypersonic Airbreathing Vehicle*, Bachelor of Engineering Thesis, The University of Queensland, Brisbane, Australia, 2009.

[22] Chobotov, V. A., *Orbital Mechanics*, American Institute of Aeronautics and Astronautics, Reston, VA, 2002.

Optimisation of Scramjet-Assisted Access-to-Space Vehicles Using Oxygen Enrichment

Sarah A. Razzaqi, Thomas Jazra, Michael K. Smart and Thomas Ferguson

Centre for Hypersonics
School of Mechanical & Mining Engineering
The University of Queensland
Brisbane, QLD 4072

Summary: The option of scramjet-assisted access to space is a key focus at The University of Queensland's Centre for Hypersonics. A three-stage, rocket-scramjet-rocket configuration is envisaged for fulfilment of this task, in which the flight envelope of the airbreathing second stage dictates the performance requirements of the complementary rocket stages. The payload mass fraction of the launch vehicle is strongly coupled with the performance of the airbreather. The present paper investigates oxygen enrichment as an option to extend the flight envelope of the second stage and thus increase the efficiency of the launch vehicle. Three Mach 12 REST scramjets are integrated with the airframe of a NASA winged-cone vehicle to assemble the second vehicle stage. Using multidisciplinary optimisation of this airbreather, the effect on the design and performance of the vehicle from supplementing the freestream oxygen flow by 10% and 20%, through the premixng of oxygen with fuel, is analysed. As suggested by a related study, the freestream dynamic pressure is held constant at q_∞ = 100 kPa during the ascent of the airbreather. It is shown that within the current vehicle model, oxygen enrichment increases the final Mach number and altitude of the second stage airbreather, which results in a larger absolute payload mass. However, the masses of airbreather and first stage rocket are also greater for the vehicles using oxygen enrichment. As a result, the payload mass fraction is only marginally different to the *baseline* non-enriched vehicle. The current model is, however, limiting in several respects, including the assumption of a constant onboard oxygen mass and the coupling of critical optimisation parameters. With refinement of the model, it is plausible that oxygen enrichment can be used to further improve vehicle performance.

I. Introduction

At present, rocket-powered launch systems are the sole means by which access to space is achievable. As these systems have been under continuous development for several decades, they are operated close to theoretical limits and only marginal further advance is achievable [1, 2]. In order to reduce costs and improve reliability, next generation launch vehicles are expected to include hypersonic airbreathing engines [3]. Scramjets in particular have been envisaged as an option for access to space using multi-stage configurations since the 1960s [4]. More recent work involving airbreather-assisted launch vehicles includes the studies of Kimura and Sawada [5], Bradford et al. [6, 7], Mehta and Bowles [8], Smart and Tetlow [9], and Tetlow and Doolan [10, 11].

In contrast to conventional rockets, scramjet-powered systems must carry only fuel onboard the vehicle since atmospheric oxygen is utilised for combustion. Scramjets have an advantage over liquid or solid-fuelled rockets in terms of a significantly higher specific impulse [12]. However, scramjet operation requires an engine-airframe integrated vehicle design [13, 14], which is inherently more complex than a "stand-alone" rocket propulsion system. Therefore, a comparison of the specific impulse in isolation distorts the picture to some extent. In fact, engine integration partially negates the advantage in specific impulse that scramjets have over rockets [15]. Nonetheless, lowering the launch costs of future space applications through the use of airbreathing propulsion is deemed possible by some [16, 17].

Effective engine integration is particularly important at hypervelocity conditions (velocity in excess of $\sim$ 3 km/s [18]) where the net thrust of the vehicle sharply decreases. There are three primary reasons for this. Firstly, as flight speeds increase, longer lengths are required to achieve adequate mixing of fuel and air in the combustor. However, longer combustors are undesirable because of the associated increase in internal drag and heat loads, so in most instances it is not practical to operate high speed scramjets near stoichiometric combustion levels. Secondly, the energy released from the combustion of hydrogen fuel becomes a smaller proportion of the specific kinetic energy of the airstream as flight speed increases. For example, at Mach 10, the ratio of available chemical energy to the specific kinetic energy of the freestream is only $\sim$ 0.77. These two factors combine to result in decreased scramjet thrust with increasing flight speed. In addition, owing to the principle of Mach number independence [18], for a given vehicle shape the external drag coefficient remains approximately constant in the hypervelocity regime. As a consequence of the decreased engine thrust and essentially constant drag, the vehicle net thrust decreases with increasing flight speed.

A thrust augmentation technique called *oxygen enrichment*, which involves the premixing of oxygen with fuel prior to injection into the engine, has been investigated for use in the hypervelocity regime [19]. Premixing oxygen with fuel provides a "head start" to the mixing process, which mitigates the poor fuel-air mixing experienced at high Mach number. The addition of onboard oxygen also allows for the combustion of fuel at greater than stoichiometric proportions (relative to the airstream). However, the benefits of oxygen enrichment come at the cost of increased propellant mass, and hence the usefulness of oxygen enrichment in scramjets must ultimately be evaluated in the context of the entire vehicle and mission.

This paper describes a preliminary study of the effects of oxygen enrichment on the design of a transatmospheric scramjet accelerator. A multidisciplinary optimisation tool for the design of hypersonic airbreathing vehicles under development at The University of Queensland [20] is used to gauge the impact of oxygen enrichment on vehicle performance. As such, a *baseline* accelerator design is optimised including the effects of enrichment, and compared to the optimal design for the non-enriched case. The following section discusses the concept of oxygen enrichment in more detail, before a description of the design and optimisation tool is provided in Sec. III. The reference vehicle and propulsion system under consideration are introduced in Secs. IV and V, respectively. Finally, the results of the study are presented along with a discussion of the benefits and drawbacks from using oxygen enrichment in Sec. VI.

II. Oxygen Enrichment

In the most general sense, oxygen enrichment can be viewed as a means to supplement the freestream oxygen flow with onboard oxygen. This mitigates the decrease in captured freestream oxygen experienced along an accelerating, constant dynamic pressure trajectory and improves engine thrust. By premixing the onboard oxygen with fuel, a further increase in thrust is achievable since the fuel/oxygen mixture is closer to a stoichiometric ratio.

A cycle analysis based study of oxygen enrichment by Rudakov and Krjutchenko [21], involving a hydrogen fuelled scramjet examined from Mach 8 to 20, found that an enriched scramjet is able to generate increased thrust while maintaining specific impulse, despite the significantly increased fuel/oxidiser flow rates. A basic vehicle study by Pike [22] (based on similarity arguments) confirms these results, indicating that from a cycle perspective, oxygen enrichment can generate the extra thrust needed at high Mach number with little degradation of specific impulse.

More recently, an experimental study of oxygen enrichment at hypervelocity conditions was conducted and further demonstrates that the technique is a viable candidate for thrust augmentation at high Mach number [19]. The study indicates that oxygen enrichment decreases required mixing lengths and increases combustion efficiency in a manner approximately proportional to the amount of premixed oxygen. The reduction in required mixing length from enrichment results in the ability to utilise a shorter combustor, which leads to reduced internal drag and heat loads, while the increase in combustion efficiency results in increased thrust.

In the current study, only the increase in combustion efficiency from oxygen enrichment is accounted for in the engine performance estimates. As such, thrust is increased for the enriched scramjet, but drag and heat loads are similar to the non-enriched scramjet. The result is an increase in scramjet net thrust for the enriched case, however, a further increase would be achievable if the mixing length reduction from enrichment was also considered.

III. Vehicle Design & Optimisation System

Several software modules are used to account for the effects of aerodynamic, propulsive, and gravitational forces during the optimisation process. The creation of the vehicle geometry and the analysis of all external aerodynamic forces and moments acting on the vehicle are undertaken using *HYPAERO* [23]. The combustion characteristics of the scramjet engines are evaluated based on quasi-one-dimensional combustor flow calculations using *dm_cycle* [24]. A modified version of the Computer Aided Design of Aerospace Concepts [25] (*CADAC*) is applied to calculate the ascent trajectory. And the Nelder/Mead [26] simplex algorithm is employed to steer all designs in the direction of the maximum Mach number upon scramjet shutdown. A summary of the above codes follows, along with an illustration of the optimisation procedure (Fig. 1).

The aerodynamics code *HYPAERO* is a preliminary design tool for the study of hypersonic vehicle shapes at low computational expense, including a meaningful prediction of viscous forces. Using strip theory, the code applies panel methods to compute pressure and Mach

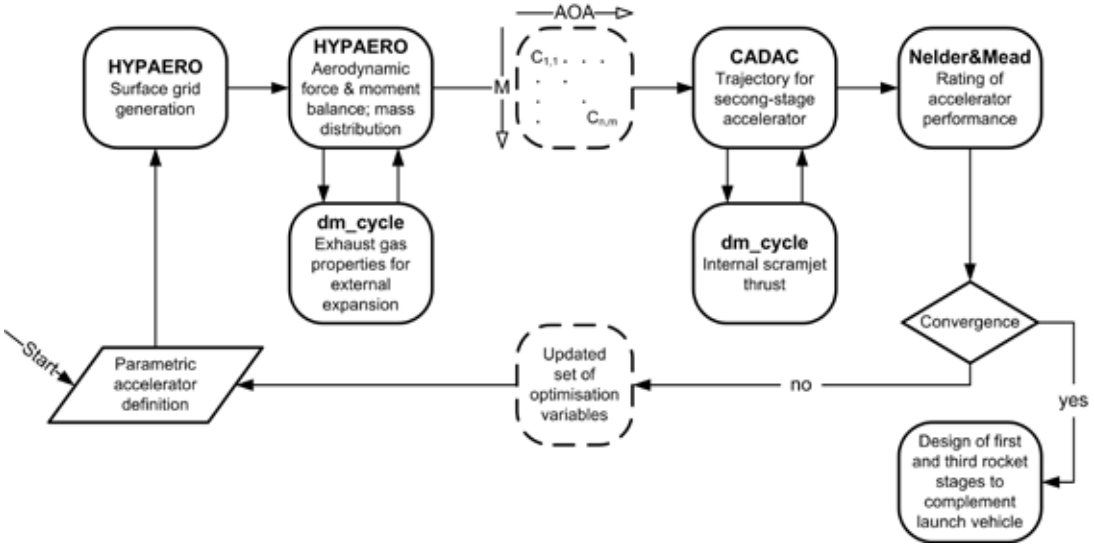


Fig. 1: Interaction of software modules during the optimisation of the accelerator.

number distributions on the vehicle surface; shear stresses are assessed by means of engineering prediction methods [27, 28], based on local flow conditions. *HYPAERO* stores essential aerodynamic vehicle properties for a range of Mach numbers and angles of attack, and also contains a model to estimate the vehicle mass. The external aerodynamics are computed separately for each vehicle component, i.e. fuselage, wings, and engine modules, and then superimposed to determine the net aerodynamic loading on the aircraft. The accuracy of the aerodynamic forces predicted by *HYPAERO* was assessed by comparing the computed lift and drag coefficients for the NASA winged-cone vehicle [29] with those obtained from a separate code called *APAS* [29], which has been deemed "acceptable for use at the conceptual design stage" [30], and wind tunnel data [31]. At Mach 10, *HYPAERO* overpredicts the lift coefficient and underpredicts the drag coefficient measured during wind tunnels tests by $\sim$ 26%. The performance of *HYPAERO* is comparable to that of *APAS* and is therefore deemed acceptable for early design stages.

Scramjet performance is evaluated using the quasi-1-D cycle analysis code *dm_cycle* [24]. Cycle analysis is an engineering approach to the modelling of supersonic combustion, based on stream thrust averaging of the engine and/or combustor flow field [32, 33]. Combustion is simulated by adding heat over a specified length inside the flow field. In *dm_cycle*, the combustor flow is computed to be a mixture of thermally perfect gases that are in thermodynamic equilibrium. For the purposes of this study, the exhaust-pressure forces obtained from the cycle analysis are incorporated into the aerodynamic coefficients in *HYPAERO*, before these are passed on to the trajectory module.

The vehicle flight path is calculated using *CADAC*, a set of software modules for the numerical simulation of spacecraft flight dynamics. It computes trajectories by integrating the equations of motion for specified time steps. At each integration time step, the current aerodynamic (from *HYPAERO*) and propulsive (from *dm_cycle*) vehicle characteristics are interpolated according to the current values of flight Mach number and angle of attack. Applying a spherical, rotating Earth model and the International Standard Atmosphere [34], a 3-degree-of-freedom (3-DOF) ascent trajectory is calculated to determine the shape-specific final Mach number of each configuration.

The non-linear, zeroth-order simplex algorithm by Nelder and Mead is used to control the optimisation process. The method has been reported to work reliably [35] and has been applied to super- and hypersonic flow problems previously [36, 13].

An engineering approach to the modelling of LH_2/LOX-fuelled first and third rocket stages by Ferguson and Jazra [37] is used to complement the rocket-scramjet-rocket launch architecture. The design approach involves minimising the fuel mass required by stages I and III respectively while catering for the predetermined flight envelope of the airbreathing second stage. As such, the first stage is sized to place the launch vehicle at the point of scramjet ignition; i.e. an altitude of 22.0 km and a flight Mach number of 6 (corresponding to a freestream dynamic pressure of 100 kPa), and a flight path angle of 0°. Based on the scramjet-shutdown conditions of the second-stage accelerator, a brief unpowered flight period to further gain altitude is modelled before the third stage is released to provide the final impulse for orbital insertion of the payload.

IV. Airbreathing Second Stage

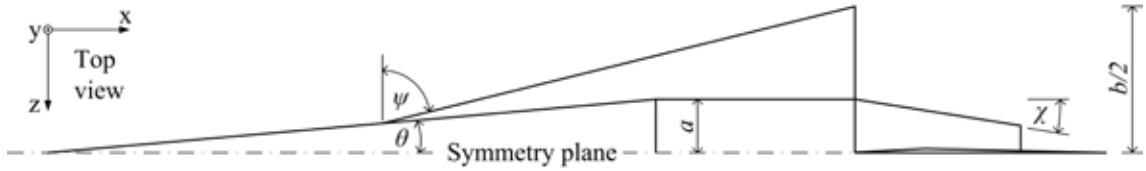
The design of the airbreathing second stage is derived from a winged-cone vehicle (WCV) originally developed at NASA with the intent to study single-stage-to-orbit vehicle design [29]. The configuration integrates a modified WCV airframe with three Mach 12 REST (rectangular-to-elliptical-shape transition) scramjet engines [38, 24]. The WCV concept is chosen as the *baseline* design for the second stage since it offers low aerodynamic drag and considerable internal volume for storage, while retaining essential characteristics of a hypersonic accelerator [33]. Figs. 2(a) and 2(b) show the design of the second-stage accelerator in the parameterised form used for this study. Seven optimisation variables have been assigned, including fuselage, wing, and scramjet parameters; the inlet capture width, w_{cap}, is not shown. Comprehensive detail on the engine design is presented in the works of Suraweera and Smart [39], and Doherty [40].

Scramjet ignition is simulated to occur at Mach 6 at an altitude of 22.0 km, which corresponds to a freestream dynamic pressure of approximately 100 kPa. This q_∞-level is maintained to within $\pm 10\%$ during the acceleration period. Scramjet shutdown is determined by the time required to burn the entire mass of the onboard liquid hydrogen. This mass is 1791 kg in the *baseline* design (see Fig. 2(c) for detail of the tank size).

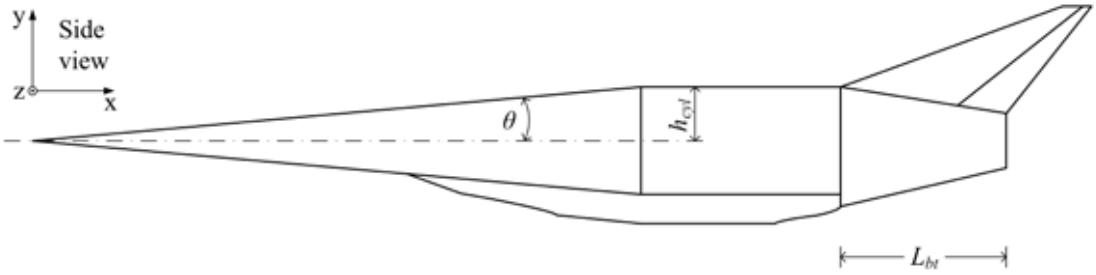
V. Propulsion System

The applicability of scramjet propulsion to access-to-space vehicles is dependent on the Mach number range over which the engines can operate. The employed scramjet engine is designed for Mach 12 flight conditions, but with a targeted operational range down to Mach 6. This engine is considered to be a near-term engine that could be envisaged to fly within the next 5-10 years. Preliminary ground testing of this engine at off-design conditions has shown promising results [39].

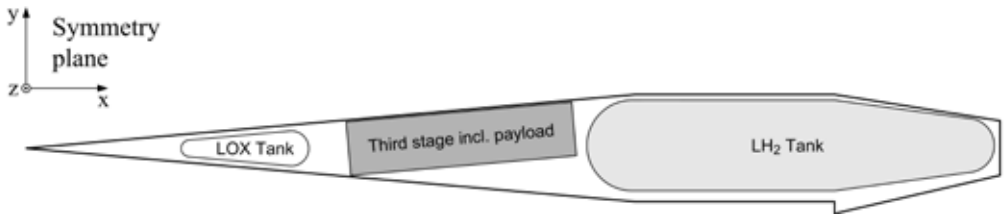
In the current study, the engine performance characteristics are determined from the quasi-one-dimensional cycle analysis tool *dm_cycle*. Four different fuelling/enrichment cases are



(a) WCV airframe



(b) Scramjet integration



(c) Internal packaging

Fig. 2: Schematic views of the scramjet-powered second stage.

considered, which are described primarily by fuel equivalence ratio, enrichment percentage, and combustion efficiency, as defined below.

The amount of fuel injected into the engine is defined relative to the freestream air only using the fuel equivalence ratio ϕ. Assuming standard air, the mass fraction of oxygen in air is 0.232. The equivalence ratio can be rewritten as

$$\phi = 8\frac{\dot{m}_{fuel}}{0.232\dot{m}_{air}} \tag{1}$$

where $\dot{m}_{fuel}$ is the mass flow rate of injected hydrogen and $\dot{m}_{air}$ is the mass flow rate of captured air.

In this work, the amount of premixed oxygen is given by the enrichment percentage, defined as

$$EP = \frac{1}{8}\frac{\dot{m}_{oxy}}{\dot{m}_{fuel}}100\% \tag{2}$$

where $\dot{m}_{oxy}$ is the mass flow rate of the premixed oxygen. In words, EP is the percentage of fuel which can be burnt by the premixed oxygen assuming complete combustion.

The combustion efficiency is defined as the percentage of injected fuel that has burnt to completion. It is specified as an input to *dm_cycle*, and determines the heat release along the combustor. Results from the experimental study of Razzaqi and Smart [19] indicate that the

Table 1: Scramjet fuelling and enrichment conditions for $M > 10$.

	ϕ [-]	EP [%]	$\eta_{c,f}$ [%]
baseline	1.0	0	80
RICH-1	1.0	10	90
RICH-2	1.0	20	100
RICH-3	1.2	20	100

increase from enrichment in the final combustion efficiency (i.e. at the end of the combustor) $\eta_{c,f}$ is approximately equal to the enrichment percentage.

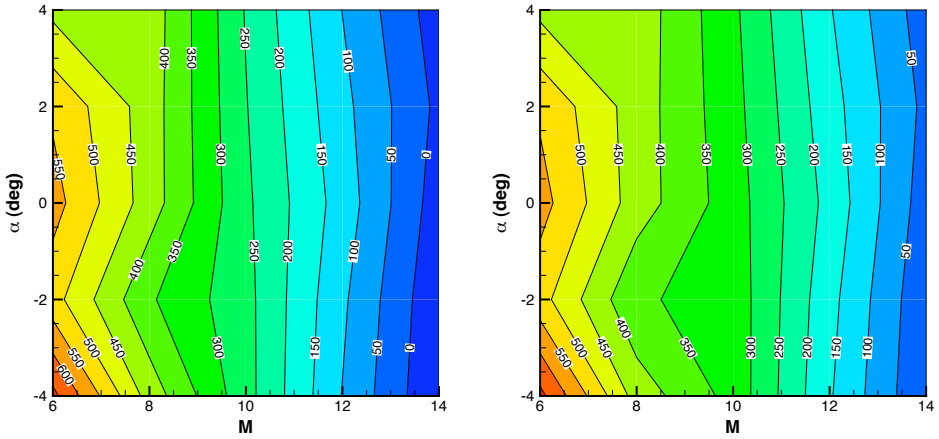
The fuelling and enrichment conditions are varied to create four different engine performance modules for inclusion in the vehicle optimisation system. From $6 < M < 10$, oxygen enrichment is not employed and the propulsion system characteristics are identical for all four cases. Above Mach 10, enrichment is employed for three of the four cases as described in Table 1.

The *baseline*, *RICH-1*, and *RICH-2* cases inject the same amount of fuel for a given mass flow of captured air (i.e. the equivalence ratio is the same), but the latter two employ enrichment at EP = 10%, 20% respectively resulting in a corresponding increase in $\eta_{c,f}$. The *RICH-3* case has the same enrichment percentage and thus combustion efficiency as *RICH-2*, however, the fuel equivalence ratio is increased to ϕ = 1.2. While injecting fuel at greater than stoichiometric proportions is a common means to increase thrust and cool the engine [32], oxygen enrichment allows for the combustion of the extra fuel with the premixed oxygen thereby generating even greater thrust.

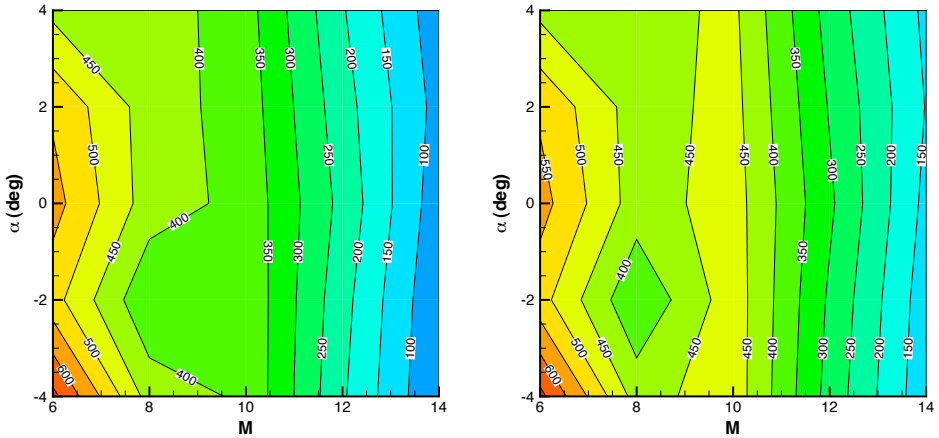
The difference in engine performance between the four cases is illustrated by the contour plots of specific thrust shown in Fig. 3. For the non-enriched case, the engine specific thrust takes a very small or even negative value in the hypervelocity regime. The use of enrichment increases specific thrust by two to three times for *RICH-3* and allows for the internal scramjet thrust to remain positive up to Mach 14. The enriched vehicles must carry the additional oxygen onboard, the mass of which is assumed to be constant for a given EP. The mass of onboard oxygen is set at 600 kg for *RICH-1* and 1200 kg for *RICH-2* and *RICH-3*. The impact of the onboard oxygen mass and of the increased thrust from enrichment is discussed in the following section.

VI. Results and Discussion

The vehicle optimisation procedure is carried out for the four fuelling cases described with the aim of maximising the Mach number at engine shutdown. In the current model, the total mass of the third stage is fixed at 1000 kg, therefore, the payload mass can be increased by decreasing the amount of propellant required by the third stage to insert the payload into the target orbit. A higher impulse of the second stage will lower the impulse required from the third stage and allow for a greater payload mass to be orbited. Oxygen enrichment will



(a) *baseline* (b) *RICH-1*



(c) *RICH-2* (d) *RICH-3*

Fig. 3: Contour plots of scramjet specific thrust obtained from dm_cycle.

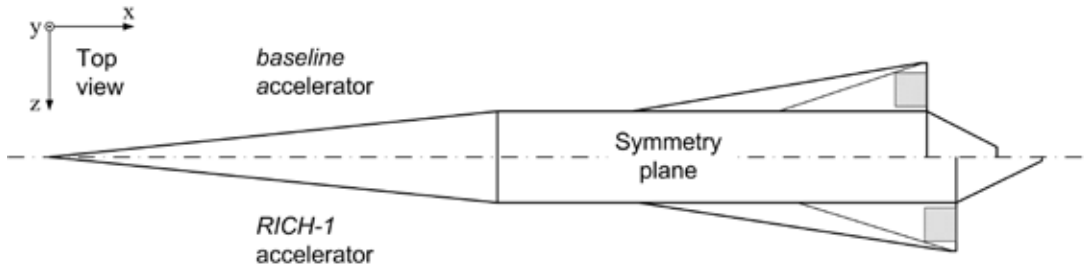
increase the airbreather shutdown Mach number, however, the increased impulse comes at the cost of an increased airbreather mass. The results presented in this section illustrate the net effect of these competing influences.

The parameters defining the geometry of the four optimised vehicles are summarised in Table 2, and presented in Fig. 4 is a schematic comparison of the *baseline* accelerator with the *RICH-1* vehicle. In general, the enriched vehicles are larger due to a slightly increased engine size in these configurations as is discussed below.

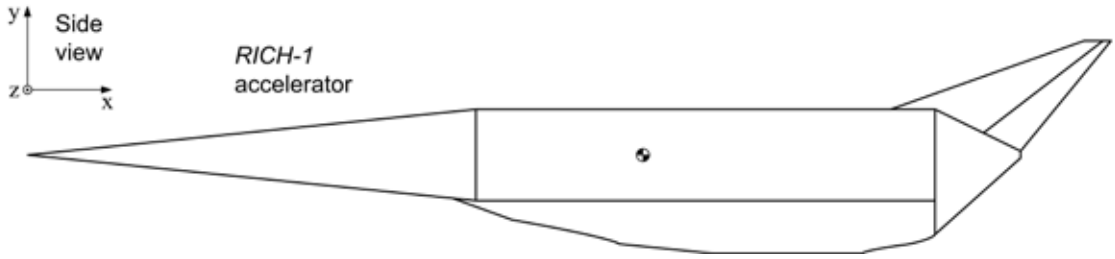
Maximising net thrust (T-D) is a critical requirement for airbreathing accelerators [41]. Within the current methodology, the optimisation system will tend to maximise the engine size, governed primarily by the desire to maximise the thrust level. However, as the engine becomes larger, the drag on the vehicle also rises. Thrust and drag do not grow proportionally, so that an optimum engine/vehicle size exists that provides for an optimum (T-D)-characteristic of the configuration. An optimised vehicle will have the largest possible engine (and hence airframe) that produces sufficient thrust to overcome the vehicle drag. The engine size is also limited by indirect mass constraints, that is, as the vehicle mass increases due to the increased engine size, the angle of attack necessary to maintain the constant dynamic pressure trajectory may fall outside the allowable range.

Table 2: Design parameters of the RICH-1, RICH-2 and RICH-3 accelerators.

	EP [%]	ϕ [-]	θ [deg]	AR [-]	χ [deg]	L_{bt} [m]	Ψ [deg]	b [m]	w_{cap} [m]
baseline	0	1.0	5.9	1.00	27.0	1.58	80.6	4.20	0.785
RICH-1	10	1.0	5.9	1.00	26.0	1.93	81.1	4.24	0.819
RICH-2	20	1.0	5.9	1.00	22.7	1.91	79.9	4.32	0.802
RICH-3	20	1.2	5.5	1.00	25.5	1.83	80.0	4.30	0.810



(a) Accelerator airframe



(b) Scramjet integration

Fig. 4: Schematic comparison of the baseline and RICH-1 accelerators.

The use of oxygen enrichment increases the specific thrust of the scramjet engine; i.e. given a particular engine size, the thrust generated using enrichment can overcome higher vehicle drag than the thrust obtained without enrichment. Thus, enrichment increases the net thrust margin allowing larger engines to be integrated with the vehicle airframe.

A secondary effect of the increased engine thrust from oxygen enrichment is the change in the boattail shape. In addition to integrating the largest possible engine, the optimisation system also tends to alter the aft-body shape to take advantage of the varied exhaust plume characteristics. As such, the engine exhaust pressure is higher when using enrichment, for instance +9% comparing the cases of *RICH-1* and the *baseline* vehicle at M = 10 and $\alpha = 2^{\circ}$. To increase the benefit from the higher exhaust-pressure level, the boattail shape of *RICH-1* has a larger projected area in the direction of thrust, approximately +18%. The effects of the increased exhaust pressure and the larger thrusting surfaces combine to produce a 39% higher pressure thrust component (the force in the direction of thrust obtained from the pressure distribution on the boattail) for *RICH-1* than the *baseline* design.

The enriched vehicles are bigger than the *baseline* vehicle, and as summarised in Table 3,

Table 3: Performance of the RICH-1, RICH-2 and RICH-3 accelerators, where h is altitude, m_{wcv} is the mass of the airbreather, m_{prop} is the mass of the airbreather propellant, m_{total} is the mass of the entire three-stage vehicle, and m_p/m_{total} is the payload mass fraction.

	EP [%]	ϕ [-]	M [-]	h [km]	m_{wcv} [kg]	m_{prop} [kg]	m_{total} [kg]	m_p/m_{total} [%]
baseline	0	1.0	11.1	30.2	7524	1791	16193	0.88
RICH-1	10	1.0	11.4	30.6	8762	2528	18800	0.85
RICH-2	20	1.0	11.7	30.9	9289	3080	19834	0.85
RICH-3	20	1.2	11.9	31.2	9243	3040	19721	0.89

they reach a higher shutdown Mach number and altitude, though at the expense of a significantly increased vehicle mass. This is largely due to the mass of the onboard oxygen and causes the requirement of substantially more powerful, heavier first-stage rockets. As a result, the payload mass fractions are very similar between all four vehicles, despite the deviations in the final Mach number and flight altitude. The *RICH-3* vehicle seemingly performs the best, however, the differences in the payload mass fractions are marginal and additional study will be required in the future to allow a more conclusive result.

One of the key constraints in the optimisation process was the assumption of a constant onboard oxygen mass (dependent only on EP). The assumption of a varying oxygen mass may cause a different angle of attack schedule than the constant mass values in this study, with implications at $M > 10$ for both thrust and drag. Also, the coupling of boattail shape and tank volume (Fig. 2(c)) constrains the fuel mass to some extent. The mass of the propellant (fuel and onboard oxygen) is critically important to the performance of the vehicle and should be included as an independent parameter in future accelerator models. It is necessary to refine the vehicle optimisation procedure, in particular concerning the above shortcomings, in order to more accurately ascertain what effect the increased propellant mass has on the vehicle performance.

The current study has been able to quantify (within the methodology used) the increase in the shutdown Mach number of a scramjet-powered accelerator available from oxygen enrichment, and also the mass penalty incurred due to the onboard storage of oxygen. In addition, first and third rocket stages were sized to calculate the payload mass fraction of three-stage launch vehicles using the optimised accelerators as the second stage. For the current system and assumptions, the benefit from the increased impulse of the enriched second-stage accelerators is offset by the additional mass required to achieve this gain. However, the restrictions placed on accelerator design and propellant mass somewhat inhibit the vehicle performance, and future enhancement of this performance seems plausible through model refinement.

In summary, this study highlights two central aspects about the nature of oxygen enrichment in scramjet-powered vehicles; firstly, potential benefits can only be appraised by an analysis of the complete mission profile of a particular vehicle design; and secondly, the key to the effective use of enrichment is an optimisation of the mass of oxygen carried onboard the vehicle.

VII. Conclusions

A preliminary vehicle optimisation study was carried out to evaluate the impact of oxygen enrichment on the design of a transatmospheric accelerator based on the NASA winged-cone vehicle and powered by three Mach 12 REST scramjets. Oxygen enrichment is applied at enrichment percentages equal to 10% and 20% and equivalence ratios of 1.0 and 1.2. Compared to the *baseline* vehicle, the enriched vehicles are bigger and heavier, but reach at best approximately 7% higher Mach number at scramjet shutdown. However, for use in rocket-scramjet-rocket launch vehicles the greater mass of the accelerator offsets its enhanced performance, such that the payload mass fractions differ only marginally between the *baseline* and the enriched configurations. This is in part a result of the constraints and assumptions placed on the implementation of oxygen enrichment in the current study, and with further refinement of the modelling the performance of the vehicle can be improved.

References

[1] Mehta, U. B., "Strategy for Developing Air-Breathing Aerospace Planes," *Journal of Aircraft*, Vol. 33, No. 2, 1996, pp. 377–385.

[2] Escher, W. J. D., editor, *The Synerjet Engine: Airbreathing/Rocket Combined-Cycle Propulsion for Tomorrow's Space Transports, Selected Papers: 1963-1996*, Society of Automotive Engineers, 1997.

[3] Cook, S. and Hueter, U., "NASA's Integrated Space Transportation Plan - 3rd Generation Reusable Launch Vehicle Technology Update," *Acta Astronautica*, Vol. 53, 2003, pp. 719–728.

[4] McClinton, C. R., Andrews, E. H., and Hunt, J. L., "Engine Development for Space Access: Past, Present and Future," *15th International Symposium on Air Breathing Engines, Bangalore, India*, 2001.

[5] Kimura, T. and Sawada, K., "Three-Stage Launch System with Scramjets," *Journal of Spacecraft and Rockets*, Vol. 36, No. 5, 1999, pp. 675–680.

[6] Bradford, J. E., Olds, J. R., and Wallace, J. G., "A Two-Stage Spacelift Concept Using Air-Breathing Propulsion and Falcon-Derived Hardware," Contractor report, SpaceWorks Engineering, Inc., 2007.

[7] Bradford, J. E., Charania, A. C., Wallace, J. G., and Eklund, D. R., "A Two-Stage to Orbit Reusable Launch Vehicle Utilizing Air-Breathing Propulsion for Responsive Space Access," AIAA Paper 2004-5950, 2004.

[8] Mehta, U. B. and Bowles, J. V., "Two-Stage-to-Orbit Spaceplane Concept with Growth Potential," *Journal of Propulsion and Power*, Vol. 17, No. 6, 2001, pp. 1149–1161.

[9] Smart, M. K. and Tetlow, M. R., "Orbital Delivery of Small Payloads Using Hypersonic Airbreathing Propulsion," *Journal of Spacecraft and Rockets*, Vol. 46, No. 1, 2009, pp. 117–125.

[10] Tetlow, M. R. and Doolan, C. J., "Orbital Payload Delivery Using Hydrogen and Hydrocarbon Fuelled Scramjet Engines," *IEEE Aerospace Conference, Big Sky, Montana*, 2006.

[11] Tetlow, M. R. and Doolan, C. J., "Launch Vehicle Mass Reduction Using a Scramjet Powered Flight Segment," *5th Australian Space Science Conference, Melbourne, Australia*, 2005.

[12] Kerrebrock, J. K., *Aircraft Engines and Gas Turbines*, The MIT Press, London, 1992.

[13] Javaid, K. H. and Serghides, V. C., "Airframe-Propulsion Integration Methodology for Waverider-Derived Hypersonic Cruise Aircraft Design Concepts," *Journal of Spacecraft and Rockets*, Vol. 42, No. 4, 2005, pp. 663–671.

[14] Wadawadigi, G., Tannehill, J. C., and Lawrence, S. L., "Three-dimensional Computation of the Integrated Aerodynamic and Propulsive Flowfields of a Generic Hypersonic Space Plane," AIAA Paper 1994-0633, 1994.

[15] Lewis, M. J., "Application of Waverider-Based Configurations to Hypersonic Vehicle Designs," AIAA Paper 1991-3304, 1991.

[16] Billig, F. S., "Research on Supersonic Combustion," *Journal of Propulsion and Power*, Vol. 9, No. 4, 1991, pp. 499-513.

[17] Curran, E. T., Heiser, W. H., and Pratt, D. T., "Fluid Phenomena in Scramjet Combustion Systems," *Annual Review of Fluid Mechanics*, Vol. 28, 1996, pp. 323–360.

[18] Anderson, J. D., *Hypersonic and High Temperature Gas Dynamics*, American Institute of Aeronautics and Astronautics, Reston, Virginia, 2006.

[19] Razzaqi, S. A. and Smart, M. K., "Shock Tunnel Experiments on Oxygen Enrichment in a Hydrogen Fuelled Scramjet," AIAA Paper 2009-7529, 2009.

[20] Jazra, T. and Smart, M. K., "Comparison of Two Optimisation Approaches for the Design of Hypersonic Airbreathing Accelerators," *Proceedings of the 9th Australian Space Science Conference*, Sydney, Australia, 2009, pp. 163–174.

[21] Rudakov, A. S. and Krjutchenko, V. V., "Additional Fuel Component Application for Hydrogen Scramjet Boosting," *SAE Aerospace Atlantic Conference & Exposition, Dayton, Ohio*, April 1990.

[22] Pike, J., "The Choice of Propellants: A Similarity Analysis of Scramjet Second Stages," *Philosophical Transactions of the Royal Society A*, Vol. 357, 1999, pp. 2357–2378.

[23] Jazra, T. and Smart, M. K., "Development of an Aerodynamics Code for the Optimisation of Hypersonic Vehicles," AIAA Paper 2009-1475, 2009.
[24] Smart, M. K., "Scramjets," *Aeronautical Journal*, Vol. 111, No. 1124, 2007, pp. 605–619.
[25] Zipfel, P. H., *Modeling and Simulation of Aerospace Vehicle Dynamics*, American Institute of Aeronautics and Astronautics, Reston, Virginia, 2007.
[26] Nelder, J. A. and Mead, R., "A Simplex Method for Function Minimisation," *Computer Journal*, Vol. 7, 1965, pp. 338–345.
[27] Meador, W. E. and Smart, M. K., "Reference Enthalpy Method Developed from Solutions of the Boundary-Layer Equations," *AIAA Journal*, Vol. 43, No. 1, 2005, pp. 135–139.
[28] van Driest, E. R., "The Problem of Aerodynamic Heating," *Aeronautical Engineering Review*, Vol. 15, No. 10, 1956, pp. 26–41.
[29] Shaughnessy, J. D., Pinckey, S. Z., McMinn, J. D., Cruz, C. I., and Kelley, M.-L., "Hypersonic Vehicle Simulation Model: Winged-Cone Configuration," Technical Memorandum 102610, NASA, 1990.
[30] Maughmer, M. D., Ozoroski, L., Straussfogel, D., and Long, L., "Validation of Engineering Methods for Predicting Hypersonic Vehicle Control Forces and Moments," *Journal of Guidance, Control, and Dynamics*, Vol. 16, No. 4, 1993, pp. 762–769.
[31] Phillips, W. P., Brauckmann, G. J., Micol, J. R., and Woods, W. C., "Experimental Investigation of the Aerodynamic Characteristics for a Winged-Cone Concept," AIAA Paper 1987-2484, 1987.
[32] Heiser, W. H. and Pratt, D. T., *Hypersonic Airbreathing Propulsion*, AIAA Education Series, Washington, DC, 1994.
[33] Curran, E. T. and Murthy, S. N. B., *Scramjet Propulsion*, American Institute of Aeronautics and Astronautics, Reston, Virginia, 2000.
[34] IHS ESDU International plc, "Equations for calculation of international standard atmosphere and associated off-standard atmospheres," *ESDU Performance Series*, No. ESDU 77022 in ESDU Product Issue 2008-03, IHS ESDU International plc, 2008.
[35] Lewis, R. M., "Direct Search Methods: Then and Now," NASA CR 2000-210125, 2000.
[36] Bowcutt, K. G., "Multidisciplinary Optimisation of Airbreathing Hypersonic Vehicles," *Journal of Propulsion and Power*, Vol. 17, No. 6, 2001, pp. 1184–1190.
[37] Ferguson, T. and Jazra, T., "Rocket Booster Design for Scramjet-Assisted Access-to-Space Vehicles," *10th Australian Space Science Conference, Brisbane, Australia*, 2010.
[38] Smart, M. K., "Design of Three-Dimensional Hypersonic Inlets with Rectangular-to-Elliptical Shape Transition," *Journal of Propulsion and Power*, Vol. 15, No. 3, 1999, pp. 408–416.
[39] Suraweera, M. V. and Smart, M. K., "Shock-Tunnel Experiments with a Mach 12 Rectangular-to-Elliptical Shape-Transition Scramjet at Off-Design Conditions," *Journal of Propulsion and Power*, Vol. 25, No. 3, 2009, pp. 555–564.
[40] Doherty, L., *Investigation of Thrust Generation in 3D Scramjets for Access-to-Space Applications*, Confirmation of Ph.D.-Candidature Report, The University of Queensland, Brisbane, Australia, 2009.
[41] Townend, L. H., "Research and Design for Hypersonic Aircraft," *Philosophical Transactions: Physical Sciences and Engineering*, Vol. 335, No. 1637, 1991, pp. 201–224.

Nonlinear Robust Control design for Hypersonic Flight Vehicles

Obaid Ur Rehman*, Ian R. Petersen*,

* *University of New South Wales at the Australian Defence Force Academy, Canberra, Australia.*

Summary: The nonlinear longitudinal equations of motion for air-breathing hypersonic flight vehicles (AHFVs) are characterized by a coupling due to the control surfaces as well as a significant coupling between engine and flight dynamics. In addition to these couplings, the non-standard dynamic characteristics of AHFVs and uncertainties in the parameter values make the flight control design of these vehicles highly complex. In the presence of input coupling terms and uncertainties, nonlinear control design using feedback linearization may not be possible. The reason for this is that the low order derivatives of the regulated outputs of the AHFV model contain an expression depending on the input, and consequently this makes feedback linearization a highly challenging task. In this paper, we apply a methodology of using a feedback linearization based uncertainty modeling approach in order to design a nonlinear robust controller for an AHFV. We consider a recent curve fitted model (CFM) model of an AHFV recently reported in the literature. The main objectives achieved in this paper are the derivation of a control-oriented linearized uncertainty model based on the CFM and the design of a nonlinear robust tracking controller to track velocity and altitude step reference commands. The linearized uncertainty model is developed by considering uncertain aerodynamic and inertial parameters. We assume that the uncertainties arising due to the couplings and parameters satisfy an integral quadratic constraint condition. This approach to modeling allow us to obtain a full vector relative degree with respect to the regulated outputs of velocity and altitude, which is one of the conditions needed to be satisfied for feedback linearization. Finally, a robust nonlinear tracking controller is constructed by using a minimax LQR approach to the tracking of the AHFV velocity and altitude. Simulation results using a curve fitted model (CFM) and varying uncertain parameters are presented to demonstrate the effectiveness of the derived linearized uncertainty model and minimax control design method.

Keywords: Feedback linearization, Robust control, Uncertainty modeling, Nonlinear systems, Hypersonic flight vehicle control.

Introduction

The nonlinear longitudinal equations of motion for air-breathing hypersonic flight vehicles (AHFVs) [1] are characterized by a coupling due to the control surfaces as well as a significant coupling between engine and flight dynamics. In addition to these couplings, the non-standard dynamic characteristics of AHFVs and uncertainties in the parameter values make the flight control design of these vehicles highly complex. If the nonlinear model of an AHFV is feedback linearizable, the complexity of this problem can be reduced to some extent [2], [3]. However, in the presence of input coupling terms, feedback linearization may not be

possible. The reason is that the low order derivatives of the regulated outputs of the AHFV model contains an expression depending on the input and consequently this makes feedback linearization a highly challenging task [4], [5]. Even with the simplified model obtained by removing all the couplings as in Parker et al [4], it is difficult to feedback linearize the system in the presence of uncertainties.

Fig. 1: An artist illustration of NASA X-43 hypersonic flight vehicle.

In this paper, we apply a methodology inspired by our previous work [6]–[8]. The methodology uses a feedback linearization based uncertainty modeling approach in order to design a nonlinear robust controller for an AHFV. In our previous work, we used a simple canard-free rigid body model of an AHFV which did not contain any input coupling [2], [3]. Unlike our previous work, in this paper we consider a more recent curve fitted model (CFM) model of an AHFV recently reported in the literature [1]. The CFM contains an extra input in the form of a canard and also exhibits input couplings and flexible effects. The CFM can also be used to design linear or nonlinear controllers based on the methods recently reported in literature such as [9], [10]. However, in order to design a high performance nonlinear robust controller in the presence of uncertainties, as demonstrated in [6], [7], we require to feedback linearize the CFM. Considering velocity and altitude as the regulated outputs, the CFM possesses a vector relative degree of $[1 \quad 1]$, and as a result nine states of the system become unobservable. This is the main reason that direct feedback linearization of the CFM is not possible and therefore, we need a new strategy to solve this control problem.

The main objectives achieved in this paper are the derivation of a control-oriented linearized uncertainty model based on a novel approach to feedback linearization of the CFM and the design of a nonlinear robust tracking controller to track velocity and altitude step reference commands. The linearized uncertainty model is developed by considering twenty four uncertain aerodynamic and inertial parameters. In contrast to the approach of Parker et al [4], where the couplings were removed from the nonlinear model, we model these couplings as uncertainties with specified maximum bounds on their magnitude. We assume that the uncertainties arising due to the couplings and parameters satisfy an integral quadratic constraint condition [11]. This approach to modeling allows us to obtain a full vector relative degree with respect to the regulated outputs of velocity and altitude, which is one of the conditions needed to be satisfied for feedback linearization [12]. Based on the derived uncertainty model for the AHFV, a robust nonlinear tracking controller is constructed by using a minimax LQR approach. Finally, the performance of the proposed method is demonstrated using simulations based on the full flexible model.

Vehicle Model

In this paper, we consider the following nonlinear model for the longitudinal dynamics of an AHFV presented in [1]:

$$\begin{aligned} \dot{V} &= \frac{T\cos\alpha - D}{m} - g\sin\gamma, \quad \dot{\gamma} = \frac{L + T\sin\alpha}{mV} - \frac{g\cos\gamma}{V}, \\ \dot{h} &= V\sin\gamma, \quad \dot{\alpha} = Q - \dot{\gamma}, \quad \dot{Q} = M_{yy}/I_{yy}, \\ \ddot{\eta}_i &= -2\zeta_m w_{m,i}\dot{\eta}_i - w_{m,i}^2\eta_i + N_i, \qquad i = 1,2,3. \end{aligned} \tag{1}$$

The nonlinear equations of motion in (1) have five rigid body states; i.e., velocity (V), altitude (h), angle of attack (α), flight path angle (γ), and pitch rate (Q). The CFM also has six vibrational modes and they are represented by generalized modal coordinates (η_i), and N_i which represent generalized forces. There are four inputs and they are the diffuser-area-ratio (A_d), throttle setting or fuel equivalence ratio (ϕ), elevator deflection (δ_e), and canard deflection (δ_c). The forces and moments in (1) do not give a closed representation of the relationship between control inputs and controlled outputs. In order to obtain tractable expressions for the purpose of control design, these forces and moments are replaced with curve-fitted approximations in [1] which is known as a curve-fitted model. The CFM has been derived by assuming a flat earth and unit vehicle depth and retains the relevant dynamical features of the actual model and also offers the advantage of being analytically tractable [1]. The approximations of the forces and moments are given as follows:

$$L \approx \bar{q}SC_L(\alpha, \delta_e, \delta_c, \Delta\tau_1, \Delta\tau_2), \tag{2}$$

$$D \approx \bar{q}SC_D(\alpha, \delta_e, \delta_c, \Delta\tau_1, \Delta\tau_2), \tag{3}$$

$$M_{yy} \approx z_T T + \bar{q}S\bar{c}C_M(\alpha, \delta_e, \delta_c, \Delta\tau_1, \Delta\tau_2), \tag{4}$$

$$T \approx \bar{q}[\phi C_{T,\phi}(\alpha, \Delta\tau_1, M_\infty) + C_T(\alpha, \Delta\tau_1, M_\infty, A_d)], \tag{5}$$

$$N_i \approx \bar{q}C_{N_i}[\alpha, \delta_e, \delta_c, \Delta\tau_1, \Delta\tau_2], \quad i = 1,2,3. \tag{6}$$

The coefficients obtained from fitting the curves are given as follows; here, we remove the function arguments for the sake of brevity:

$$C_L = C_L^\alpha\alpha + C_L^{\delta_e}\delta_e + C_L^{\delta_c}\delta_c + C_L^{\Delta\tau_1}\Delta\tau_1 + C_L^{\Delta\tau_2}\Delta\tau_2 + C_L^0, \tag{7}$$

$$C_M = C_M^\alpha\alpha + C_M^{\delta_e}\delta_e + C_M^{\delta_c}\delta_c + C_M^{\Delta\tau_1}\Delta\tau_1 + C_M^{\Delta\tau_2}\Delta\tau_2 + C_M^0, \tag{8}$$

$$\begin{aligned} C_D =& C_D^{(\alpha+\Delta\tau_1)^2}(\alpha+\Delta\tau_1)^2 + C_D^{(\alpha+\Delta\tau_1)}(\alpha+\Delta\tau_1) + C_D^{\delta_e^2}\delta_e^2 + C_D^{\delta_e}\delta_e + C_D^{\delta_c^2}\delta_c^2 + C_D^{\delta_c}\delta_c + C_D^{\alpha\delta_e}\alpha\delta_e \\ &+ C_D^{\alpha\delta_c}\alpha\delta_c + C_D^{\delta\tau_1}\delta\tau_1 + C_D^0, \end{aligned} \tag{9}$$

$$C_{T,\phi} = C_{T,\phi}^\alpha\alpha + C_{T,\phi}^{\alpha M_\infty^{-2}}\alpha M_\infty^{-2} + C_{T,\phi}^{\alpha\Delta\tau_1}\alpha\Delta\tau_1 + C_{T,\phi}^{M_\infty^{-2}}M_\infty^{-2} + C_{T,\phi}^{{\Delta\tau_1}^2}{\Delta\tau_1}^2 + C_{T,\phi}^{\Delta\tau_1}\Delta\tau_1 + C_{T,\phi}^0, \tag{10}$$

$$C_T = C_T^{A_d}A_d + C_T^\alpha\alpha + C_T^{M_\infty^{-2}}M_\infty^{-2} + C_T^{\Delta\tau_1}\Delta\tau_1 + C_T^0, \tag{11}$$

$$C_{N_i} = C_{N_i}^\alpha\alpha + C_{N_i}^{\delta_e}\delta_e + C_{N_i}^{\delta_c}\delta_c + C_{N_i}^{\Delta\tau_1}\Delta\tau_1 + C_{N_i}^{\Delta\tau_2}\Delta\tau_2 + C_{N_i}^0, \tag{12}$$

where M_∞ and $\bar{q}$ are defined as follows:

$$\bar{q} = \frac{\rho(h)V^2}{2}, \quad M_\infty = \frac{V}{M_0}. \tag{13}$$

Uncertainty modeling for robust controller synthesis

Simplification of the CFM

In the presence of the input coupling in the lift and drag coefficients of the CFM, an attempt to feedback linearize the system leads to input terms in its low order derivatives [4]. In order to obtain a form suitable for feedback linearization, we do some simplification in the model by considering some of the terms in the coefficients as uncertainty including $\Delta\tau_1,\ \Delta\tau_2$ which represent the rear and aft angle due to the structure flexibility. The simplified expressions for lift, moment, drag, and thrust coefficients now can be written as follows:

$$\begin{aligned}
C_{T,\phi} &= C_{T,\phi}^{\alpha}\alpha + C_{T,\phi}^{\alpha M_\infty^{-2}}\alpha M_\infty^{-2}\alpha + C_{T,\phi}^{M_\infty^{-2}} M_\infty^{-2} + C_{T,\phi}^{0} + \Delta C_{T,\phi},\\
C_L &= C_L^{\alpha}\alpha + C_L^0 + \Delta C_l,\\
C_M &= C_M^{\alpha}\alpha + [C_M^{\delta_e} - C_M^{\delta_c}(\frac{C_L^{\delta_e}}{C_L^{\delta_c}})]\delta_e + C_M^0 + \Delta C_M,\\
C_D &= C_D^{(\alpha+\Delta\tau_1)^2}(\alpha)^2 + C_D^{(\alpha+\Delta\tau_1)}(\alpha) + C_D^0 + \Delta C_d,\\
C_T &= C_T^{A_d}A_d + C_T^{\alpha}\alpha + C_T^{M_\infty^{-2}} M_\infty^{-2} + C_T^0 + \Delta C_T,
\end{aligned} \tag{14}$$

where ΔC_d, ΔC_l, ΔC_T, $\Delta C_{T,\phi}$ and ΔC_M represent the uncertainties in their corresponding functions. Furthermore, in order to obtain full relative degree for the purpose of feedback linearization, we dynamically extend the system by introducing second order actuator dynamics into the fuel equivalence ratio input as follows:

$$\ddot{\phi} = -2\zeta\omega_n\dot{\phi} - \omega_n^2\phi + \omega_n^2\phi_c. \tag{15}$$

After this extension we have two more states ϕ and $\dot{\phi}$, and thus the sum of the elements of vector relative degree is equal to the order of the system (n); i.e. $n = 7$ and thus satisfying one of the conditions for feedback linearization [12].

Feedback Linearization of the Simplified Model

The model obtained through the above simplification is still difficult to feedback linearize due to the presence of uncertainties in the system. We approach this problem by using the technique developed in [6]. The outputs to be regulated are selected as velocity (V) and altitude (h) using two inputs, elevator deflection (δ_e) and fuel equivalence ratio (Φ_c). Since δ_c is a function of δ_e; i.e related to δ_e via an interconnect gain, we do not consider it as a separate input. Furthermore, we fix the diffuser area ratio (A_d) to be unity. This manipulation results in a two input and two output square system. The vehicle trimmed cruise conditions are given in Table 1. The new simplified model consist of seven rigid states and two additional integral states as follows:

$$x = \begin{bmatrix} V_I & V & h_I & h & \gamma & \alpha & \phi & \dot{\phi} & Q \end{bmatrix}^T, \tag{16}$$

where,

$$V_I = \int_0^t (V(\tau) - V_c)d\tau, \quad h_I = \int_0^t (h(\tau) - h_c)d\tau,$$

and V_c and h_c are the desired command values for the velocity and altitude respectively. The uncertain parameter vector $p \in \mathbb{R}^{24}$ includes the vehicle inertial parameters, coupling

terms and the coefficients which appear in the forces and moments approximations described previously and is given as follows:

$$p = [1/m \quad \bar{c} \quad 1/I_{yy} \quad \rho \quad S \quad C_L^{\alpha} \quad C_L^0 \quad C_D^{(\alpha)^2} \quad C_{T,\phi}^{\alpha} \quad C_D^{(\alpha)} \quad C_D^0 \quad \Delta C_l \quad \Delta C_d \quad C_M^{\alpha} \quad C_M^0 \quad C_M^{\delta_c} \quad C_M^{\delta_e} \quad C_{T,\phi}^0 \quad C_T^{A_d} \quad C_T^{\alpha} \quad \Delta C_T \quad \Delta C_M \quad \Delta C_{T,\phi} \quad C_T^0]^T. \tag{17}$$

The nominal value of p is denoted by p_0 and uncertainty in p is denoted by Δp such that $p = p_0 + \Delta p$. The model of the AHFV can be written in a general multi-input multi-output (MIMO) form as follows:

$$\begin{aligned} x(t) &= f(x, p_0) + \sum_{k=1}^{2} g_k(x, p_0)u_k + \Delta f(x, p) + \sum_{k=1}^{2} \Delta g_k(x, p)u_k; \\ y_i(t) &= \nu_i(x, p), \quad i = 1, 2. \end{aligned} \tag{18}$$

where, $\Delta f(x, p)$ and $\Delta g_k(x, p)$ are the uncertainty terms appearing in the corresponding functions. The control vector u and output vector y are defined as

$$u = [u_1, u_2]^T = [\delta_e, \phi_c]^T, y = [y_1, y_2]^T = [V, h]^T.$$

We assume that the $p \in \Psi$, where Ψ is a compact convex set that represents the admissible range of variation of p such that p_0 lies in its interior. In this study, a maximum variation of 10% of the nominal values has been considered. Thus, $\Psi = \{p \in \mathbb{R}^{24} \mid |0.9p_{i0}| \leq |p_i| \leq |1.1p_{i0}| \quad \text{for} \quad i = 1, \cdots, 24\}$.

Derivation of linearized uncertain model for AHFV

For feedback linearization purposes, the selected outputs V and h are differentiated as many times as required until one of the inputs appears in the resulting expression. This can be achieved by differentiating V three times and h four times using the Lie derivative [8]. Assuming $\Delta W_{V^{(2)}}(x, p)$, $\Delta W_{h^2}(x, p)$, $\Delta W_{h^{(3)}}(x, p)$, $\Delta W_{V^{(3)}}$ and $\Delta W_{h^{(4)}}$ are the uncertainties in their respective functions, we can write differentiated outputs as sum of nominal and uncertainty parts as follows:

$$\begin{aligned} \ddot{V}(x, p) &= \underbrace{\ddot{V}(x, p_0)}_{\text{nominal part}} + \underbrace{\Delta W_{V^{(2)}}(x, u, p)}_{\text{uncertainty part}}, \\ \dddot{V}(x, p) &= \underbrace{\dddot{V}(x, p_0)}_{\text{nominal part}} + \underbrace{\Delta W_{V^{(3)}}(x, u, p)}_{\text{uncertainty part}}, \\ \ddot{h}(x, p) &= \underbrace{\ddot{h}(x, p_0)}_{\text{nominal part}} + \underbrace{\Delta W_{h^{(2)}}(x, u, p)}_{\text{uncertainty part}}, \\ \dddot{h}(x, p) &= \underbrace{\dddot{h}(x, p_0)}_{\text{nominal part}} + \underbrace{\Delta W_{h^{(3)}}(x, u, p)}_{\text{uncertainty part}}, \\ h^{(4)}(x, p) &= \underbrace{h^{(4)}(x, p_0)}_{\text{nominal part}} + \underbrace{\Delta W_{h^{(4)}}(x, u, p)}_{\text{uncertainty part}}. \end{aligned} \tag{19}$$

In order to feedback linearize the system, we can write the second and fifth equations in (19) in the following form:

$$\begin{bmatrix} \dddot{V}(x, u, p) \\ h^4(x, u, p) \end{bmatrix} \underbrace{= f_*(x, p_0) + g_*(x, p_0) \begin{bmatrix} \delta_e \\ \phi_c \end{bmatrix}}_{\text{Nominal nonlinear part}} + \underbrace{\begin{bmatrix} \Delta W_{V^{(3)}}(x, u, p) \\ \Delta W_{h^{(4)}}(x, u, p) \end{bmatrix}}_{\text{Uncertain nonlinear part}}, \tag{20}$$

where

$$f_*(x,p_0) = \begin{bmatrix} L_f^3 V \\ L_f^4 h \end{bmatrix}, \quad g_*(x,p_0) = \begin{bmatrix} L_{g_1}(L_f^2 V) & L_{g_2}(L_f^2 V) \\ L_{g_1}(L_f^3 h) & L_{g_2}(L_f^3 h) \end{bmatrix}.$$

The application of the feedback linearizing control law [8]

$$u = -g_*(x,p_0)^{-1} f_*(x,p_0) + g_*(x,p_0)^{-1} v \tag{21}$$

can only cancel the nominal nonlinear part of the equations as follows:

$$\begin{bmatrix} \dddot{V}(x,u,p) \\ h^4(x,u,p) \end{bmatrix} = \underbrace{\begin{bmatrix} v_1 \\ v_2 \end{bmatrix}}_{\text{Nominal linear part}} + \underbrace{\begin{bmatrix} \Delta W_{V^{(3)}}(x,u,p) \\ \Delta W_{h^{(4)}}(x,u,p) \end{bmatrix}}_{\text{Uncertain nonlinear part}}. \tag{22}$$

It is to be noted that in (22), the nominal part is linearized but the uncertainty part is still nonlinear in x and p. The left hand side of (22) shows the higher order derivatives of the regulated variables V and h. In order to proceed further, they are required to be written in first order derivative form. For this purpose, we define a nominal nonlinear diffeomorphism which transforms (22) into a new set of coordinates. We define new state vectors $\xi = \begin{bmatrix} \xi_1 & \xi_2 & \xi_3 & \xi_4 \end{bmatrix}^T$ and $\eta = \begin{bmatrix} \eta_1 & \eta_2 & \eta_3 & \eta_4 & \eta_5 \end{bmatrix}^T$ as follows:

$$\xi_1 = V_I, \quad \xi_2 = V - V_c, \quad \xi_3 = \dot{V}, \quad \xi_4 = \ddot{V}, \tag{23}$$

$$\eta_1 = h_I, \quad \eta_2 = h - h_c, \quad \eta_3 = \dot{h}, \quad \eta_4 = \ddot{h}, \quad \eta_5 = \dddot{h}. \tag{24}$$

We also define a local nominal nonlinear diffeomorphism $T_1(x,p_0,V_c)$, and $T_2(x,p_0,h_c)$ assuming a nominal model that which maps the new vectors ξ and η respectively to the original vector x as follows:

$$\xi = T_1(x,p_0,V_c), \quad \eta = T_2(x,p_0,h_c), \tag{25}$$

where,

$$T_1(x,p_0,V_c) = \begin{bmatrix} V_I & V - V_c & \dot{V}(x,p_0) & \ddot{V}(x,p_0) \end{bmatrix}^T,$$
$$T_2(x,p_0,h_c) = \begin{bmatrix} h_I & h - h_c & \dot{h}(x) & \ddot{h}(x,p_0) & \dddot{h}(x,p_0) \end{bmatrix}.$$

Furthermore, we can write (25) as follows:

$$\chi = T(x,p_0,V_c,h_c), \tag{26}$$

where $\chi = \begin{bmatrix} \xi & \eta \end{bmatrix}^T$, and $T(x,p_0,V_c,h_c) = \begin{bmatrix} T_1(x,p_0,V_c) & T_2(x,p_0,h_c) \end{bmatrix}^T$. In order to write the original state (x) in terms of the new state vector (χ), we find an inverse of (26) as

$$x = T^{-1}(\chi). \tag{27}$$

To find an approximation to the required inverse, we expand (26) using a first order Taylor expansion at the nominal trim conditions by assuming that the linearization is valid in a neighborhood of the trim condition as follows:

$$\chi \simeq \chi(0) + \nabla_x T(x(0))(x - x(0)), \tag{28}$$

where $\chi_0 = \begin{bmatrix} \xi(0) & \eta(0) \end{bmatrix}^T$ is the initial condition for the new state vector. Also, $\nabla_x T(x(0))$ is the gradient matrix at $x(0)$ and is defined as

$$\nabla_x T(x(0)) = \frac{\partial T(x)}{\partial x}|_{x=x(0)}, \tag{29}$$

Table 1: States, Inputs and Selected Physical Parameters at the Trim Condition

Vehicle Velocity (V)	7702.08 ft/sec
Altitude (h)	85000 ft
Fuel-to-air ratio (ϕ)	0.4388
Pitch Rate Q	0
Angle of Attack (α)	-0.0134 rad
Flight Path Angle (γ)	0
Elevator Deflection (δ_e)	2.005 deg
Canard Deflection (δ_c)	2.79 deg
Diffuser Area ratio (A_d)	1
Reference Area (S)	17 sq-ft.ft^{-1}
Mean Aerodynamic Chord (c)	17 ft
Air Density (ρ)	6.6×10^{-5} slugs/ft^3
Mass with 50% fuel level (m)	147.9 slug. ft^{-1}
Moment of Inertia (I_{yy})	5.0×10^5 slugs/ft^2/(rad . ft)

where $t_1, t_2 ..... , t_m$ are the elements of the vector $T(x)$ and $x_1, x_2 ..... x_n$ are the elements of original state vector x with $m = n = 9$:

$$\frac{\partial T(x)}{\partial x} = \begin{bmatrix} \frac{\partial t_1}{\partial x_1} & \frac{\partial t_1}{\partial x_2} & \cdots & \frac{\partial t_1}{\partial x_n} \\ \frac{\partial t_2}{\partial x_1} & \frac{\partial t_2}{\partial x_2} & \cdots & \frac{\partial t_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial t_m}{\partial x_1} & \frac{\partial t_m}{\partial x_2} & \cdots & \frac{\partial t_m}{\partial x_n} \end{bmatrix}.$$

Solving (28) for x yields the transformation equation for x:

$$x \simeq \nabla_x T^{-1}(x(0))(\chi - \chi(0)) + x(0). \tag{30}$$

Using the definition in (29), we can easily find the matrix $\nabla_x T(x(0))$. Furthermore, the mapping $u \rightarrow v$ can be derived using (21) and by transforming into the new states using (30) as follows:

$$u = -\tilde{g}(\chi, p_0)^{-1} \tilde{f}(\chi, p_0) + \tilde{g}(\chi, p_0)^{-1} v, \tag{31}$$

where $\tilde{f}(\chi, p_0)$ and $\tilde{g}(\chi, p_0)$ are the nonlinear functions corresponding to $f_*(x, p_0)$ and $g_*(x, p_0)$ in terms of the new states respectively. Using the definition of the new states, and equations (19), and assuming that no perturbations exist in $\dot{\xi}_1$, $\dot{\xi}_2$, $\dot{\eta}_1$, and $\dot{\eta}_2$ because V, $\dot{V}$, h, and all other states in x are measurable, we write the differential equations as follows:

$$\begin{bmatrix} \dot{\xi}_1 \\ \dot{\xi}_2 \\ \dot{\xi}_3 \\ \dot{\xi}_4 \\ \dot{\eta}_1 \\ \dot{\eta}_2 \\ \dot{\eta}_3 \\ \dot{\eta}_4 \\ \dot{\eta}_5 \end{bmatrix} = \begin{bmatrix} \xi_2 \\ \xi_3 \\ \xi_4 \\ v_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \\ \eta_5 \\ v_2 \end{bmatrix} + \underbrace{\begin{bmatrix} 0 \\ 0 \\ \Delta W_{V^{(2)}}(x,p) \\ \Delta W_{V^{(3)}}(x,u,p) \\ 0 \\ 0 \\ \Delta W_{h^{(2)}}(x,p) \\ \Delta W_{h^{(3)}}(x,p) \\ \Delta W_{h^{(4)}}(x,u,p) \end{bmatrix}}_{\text{Uncertain nonlinear terms}}. \tag{32}$$

In order to linearize equation (32), we use a first order Taylor expansion with respect to each uncertainty Δp_s for $s = 1, 2, \cdots, 24$ and the new states. The linearized equations in terms of the new states and new inputs can be written as follows:

$$\dot{\chi} = \begin{bmatrix} \xi_2 \\ \xi_3 \\ \xi_4 \\ v_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \\ \eta_5 \\ v_2 \end{bmatrix} + \sum_{s=1}^{24} \Delta p_s \begin{bmatrix} 0 \\ 0 \\ \Delta\tilde{w}_{V^{(2)}(p_s)} \\ \Delta\tilde{w}_{V^{(3)}(p_s)} \\ 0 \\ 0 \\ \Delta\tilde{w}_{h^{(2)}(p_s)} \\ \Delta\tilde{w}_{h^{(3)}(p_s)} \\ \Delta\tilde{w}_{h^{(4)}(p_s)} \end{bmatrix} \chi + \sum_{s=1}^{24} \Delta p_s \begin{bmatrix} 0 \\ 0 \\ 0 \\ \Delta\check{w}_{V^{(3)}(\Delta p_s)} \\ 0 \\ 0 \\ 0 \\ \Delta\check{w}_{h^{(4)}(\Delta p_s)} \end{bmatrix} v \tag{33}$$

Furthermore, to write the linearized system (33) in a suitable linear form, we define the corresponding matrices as follows:

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}; \quad B = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}; \quad C_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}; \quad K_1 = \begin{bmatrix} \Delta w_{V^{(2)}(1)} \\ \Delta w_{V^{(2)}(2)} \\ \vdots \\ \Delta w_{V^{(2)}(23)} \\ \Delta w_{V^{(2)}(24)} \end{bmatrix};$$

$$C_2 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}; \quad K_2 = \begin{bmatrix} \Delta w_{V^{(3)}(1)} \\ \Delta w_{V^{(3)}(2)} \\ \vdots \\ \Delta w_{V^{(3)}(23)} \\ \Delta w_{V^{(3)}(24)} \end{bmatrix}; C_3 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}; \quad K_3 = \begin{bmatrix} \Delta w_{h^{(2)}(1)} \\ \Delta w_{h^{(2)}(2)} \\ \vdots \\ \Delta w_{h^{(2)}(23)} \\ \Delta w_{h^{(2)}(24)} \end{bmatrix};$$

$$C_4 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}; \quad K_4 = \begin{bmatrix} \Delta w_{h^{(3)}(1)} \\ \Delta w_{h^{(3)}(2)} \\ \vdots \\ \Delta w_{h^{(3)}(23)} \\ \Delta w_{h^{(3)}(24)} \end{bmatrix}; C_5 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}; \quad K_5 = \begin{bmatrix} \Delta w_{h^{(4)}(1)} \\ \Delta w_{h^{(4)}(2)} \\ \vdots \\ \Delta w_{h^{(4)}(23)} \\ \Delta w_{h^{(4)}(24)} \end{bmatrix};$$

$$G_2 = \begin{bmatrix} \Delta\check{w}_{1(1)} \\ \Delta\check{w}_{1(2)} \\ \vdots \\ \Delta\check{w}_{1(23)} \\ \Delta\check{w}_{1(24)} \end{bmatrix}; \quad G_5 = \begin{bmatrix} \Delta\check{w}_{2(1)} \\ \Delta\check{w}_{2(2)} \\ \vdots \\ \Delta\check{w}_{2(23)} \\ \Delta\check{w}_{2(24)} \end{bmatrix}, \quad G_1 = \mathbf{0}, \quad G_3 = \mathbf{0}, \quad G_4 = \mathbf{0}, \quad \Delta = diag(\Delta p), \tag{34}$$

and write the equation (33) as:

$$\dot{\chi}(t) = [A + \sum_{j=1}^{5} C_j \Delta(t) K_j]\chi(t) + [B + \sum_{j=1}^{5} C_j \Delta(t) G_j]v(t); \quad \|\Delta\| \leq 1, \tag{35}$$

where, $\chi(t) \in \mathbb{R}^9$, $v(t) = [v_1 \ v_2]^T \in \mathbb{R}^2$ and $z_j(t) \in \mathbb{R}^{24}$, for $j = 1, 2, \cdots, 5$ (see Fig. 2). Hence, (35) can be written in a general MIMO form by defining $\zeta_j(t) = \Delta[K_j\chi(t) + G_j v(t)] \in \mathbb{R}$ for $j = 1, 2, \cdots, 5$ as follows:

$$\begin{aligned} \dot{\chi}(t) &= A\chi(t) + Bv(t) + \sum_{j=1}^{5} C_j \zeta_j(t); \\ z_j(t) &= K_j\chi(t) + G_j v(t); \text{for } j = 1, 2, \cdots, 5 \end{aligned} \tag{36}$$

where $z_j(t) \in \mathbb{R}^{24}$ are the uncertainty outputs (see Fig. 2). In this paper, the linearized model (36) is referred as control oriented linearized uncertainty model (CLUM) for the AHFV at the given nominal trim condition.

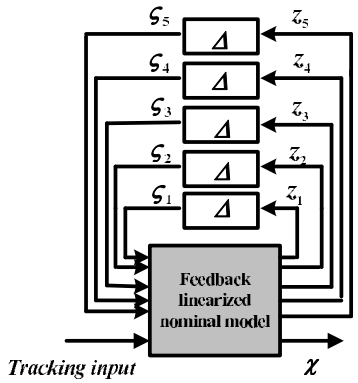


Fig. 2: Block diagram representation of feedback linearized uncertainty model for AHFV

Minimax LQR Control system design

The minimax optimal control method is an extension of the LQR optimal control method for a class of uncertain systems. A minimax optimal controller provides robust stability as well as robust performance for a system with structured uncertainty by incorporating the uncertainty bound directly into the controller synthesis. Furthermore, the minimax optimal control minimizes the maximum value of a quadratic cost function (37) and gives an optimized controller for the selected cost function. The quadratic cost function is given as follows:

$$F = \int_0^\infty [\chi(t)^T Q \chi(t) + v(t)^T R v(t)] dt, \tag{37}$$

where $Q = Q^T > 0$ and $R = R^T > 0$ are given state and control weighting matrices respectively. The minimax optimal controller can be constructed by solving a game type parameter dependent Riccati equation (38) for given parameters $\tau_j > 0$ [11]:

$$(A - BE^{-1}G^T K)^T X_\tau + X_\tau (A - BE^{-1}G^T K) + X_\tau (CC^T - BE^{-1}B^T) X_\tau + K^T (I - GE^{-1}G^T) K = 0 \tag{38}$$

where,

$$K = \begin{bmatrix} Q^{1/2} \\ 0 \\ \sqrt{\tau_1} K_1 \\ \vdots \\ \sqrt{\tau_5} K_5 \end{bmatrix}, \quad G = \begin{bmatrix} 0 \\ R^{1/2} \\ \sqrt{\tau_1} G_1 \\ \vdots \\ \sqrt{\tau_5} G_5 \end{bmatrix}, \quad E = GG^T,$$
$$C = \begin{bmatrix} \frac{1}{\sqrt{\tau_1}} C_1 & \frac{1}{\sqrt{\tau_2}} C_2 & \frac{1}{\sqrt{\tau_3}} C_3 & \frac{1}{\sqrt{\tau_4}} C_4 & \frac{1}{\sqrt{\tau_5}} C_5 \end{bmatrix}. \tag{39}$$

If X_τ, the solution of Riccati equation (38) for the given parameters τ_j, is symmetric and positive definite, the controller leads to a stable closed loop system [11]. In addition, the optimum value of the parameters are selected in such a way that they provide the minimum value of the bound (40) on the cost function (37):

$$\min[\chi^T(0) X_\tau \chi(0) + \sum_{j=1}^{5} \tau_j \chi^T(0) D_j \chi(0)], \tag{40}$$

where $D_j > 0$ for $j = 1, 2, \cdots, 5$. The minimax optimal control law, for given values of the optimized parameter τ and cost matrices Q and R, is given as follows [11]:

$$v(t) = -G_\tau \chi(t), \tag{41}$$

where G_τ is the controller gain matrix defined as

$$G_\tau = E^{-1}[B^T X_\tau + G^T K].$$

Following the procedure described above, we design a minimax optimal controller producing the control signal v by selecting appropriate weights for the system and solve the Riccati equation (38) depending on the parameters τ. The selected weights and the value of the parameters are given below:

$$\mathbf{Q} = \mathbf{diag}\left[\ 5, 10, 75, 100, 0.1, 10, 10, 100, 10\ \right], \tag{42}$$

$$\mathbf{R} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \tau_1 = 68.09,\ \tau_2 = 32.7731,\ \tau_3 = 81.83,\ \tau_4 = 57.2751,\ \tau_5 = 77.33. \tag{43}$$

The complete closed loop structure is shown in Fig (3).

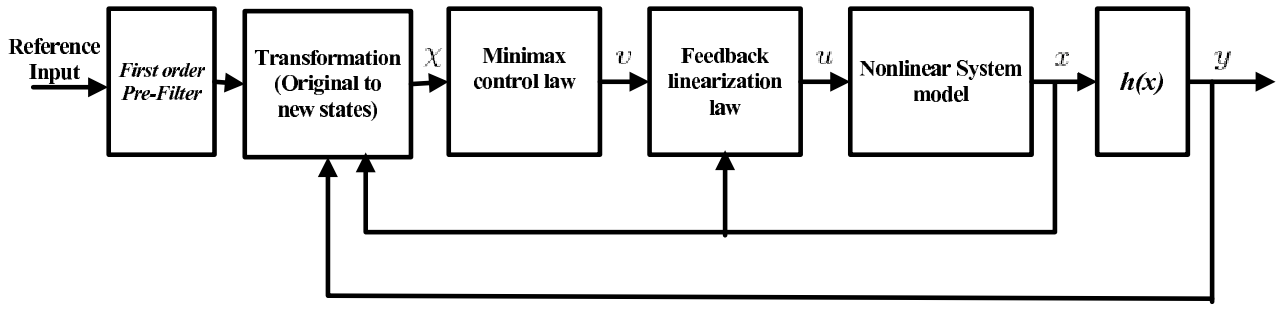


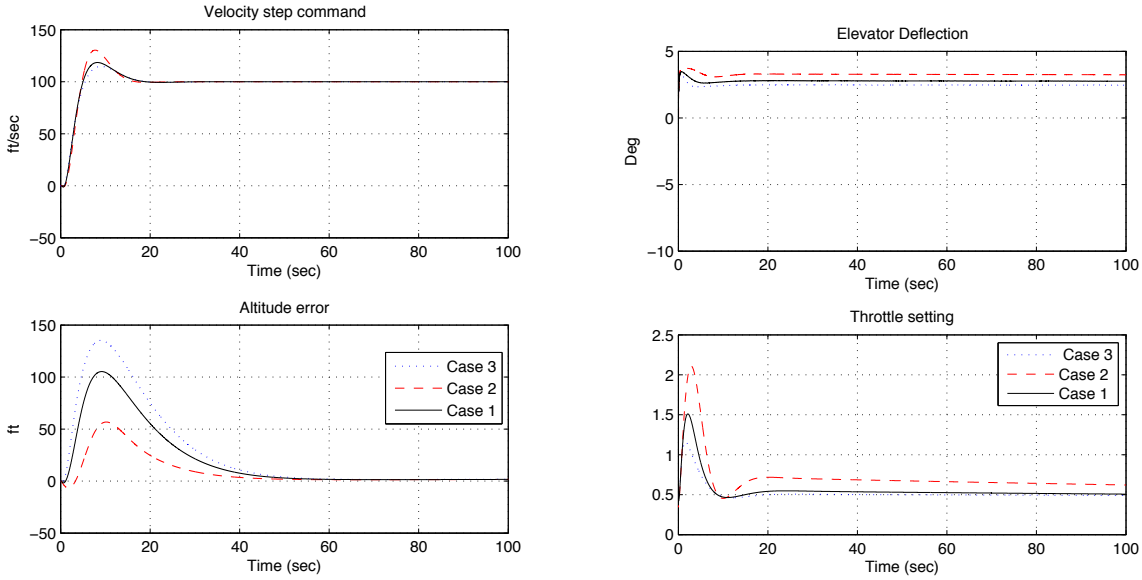
Fig. 3: Closed loop system structure for AHFV

Simulation Results

Simulation studies, for the proposed scheme have been conducted based on the actual (unsimplified) uncertain nonlinear longitudinal AHFV motion dynamics model (1) and by considering 24 parametric uncertainties. In order to analyze the performance of the proposed controller, we simulated the complete closed loop system with several cases of different combination of uncertainties. Here, for the sake of brevity the simulation results for only three cases of uncertainties are presented. In these three cases which are explained below we evaluate the performance of the proposed controller by using step input commands.

1. In a first case, we consider a nominal model i.e. without any any uncertainties.
2. In the second case, we consider the uncertainties which are 10% lower than their corresponding nominal values.
3. In the third case, we consider the uncertainties which are 10% larger than their corresponding nominal values.

The responses of the minimax optimal controller for the above combinations of uncertainties with a $100\ ft/sec$ step change in velocity are shown in Fig. 4 and with a $2000\ ft$ step change in altitude are shown in Fig. 5. The solid lines show the responses without any uncertainties,



(a) (b)

Fig. 4: (a) Response to a step change in velocity, (b) Control inputs.

large dash lines show the responses for second case and small dash lines show the responses for the third case. The results in Fig. 4 show that the controller track the velocity step command in reasonable time without leaving any steady state error. Furthermore, the controller outputs and altitude error remain low and practical in achieving the velocity tracking requirements for all the three cases described above. Similarly, the results in Fig. 5 show that the altitude tracking responses are also very good. In summary, we can say that the proposed robust tracking controller gives good results in terms of robustness and performance in the nominal case as well as in the cases of uncertainties and utilizes low control effort.

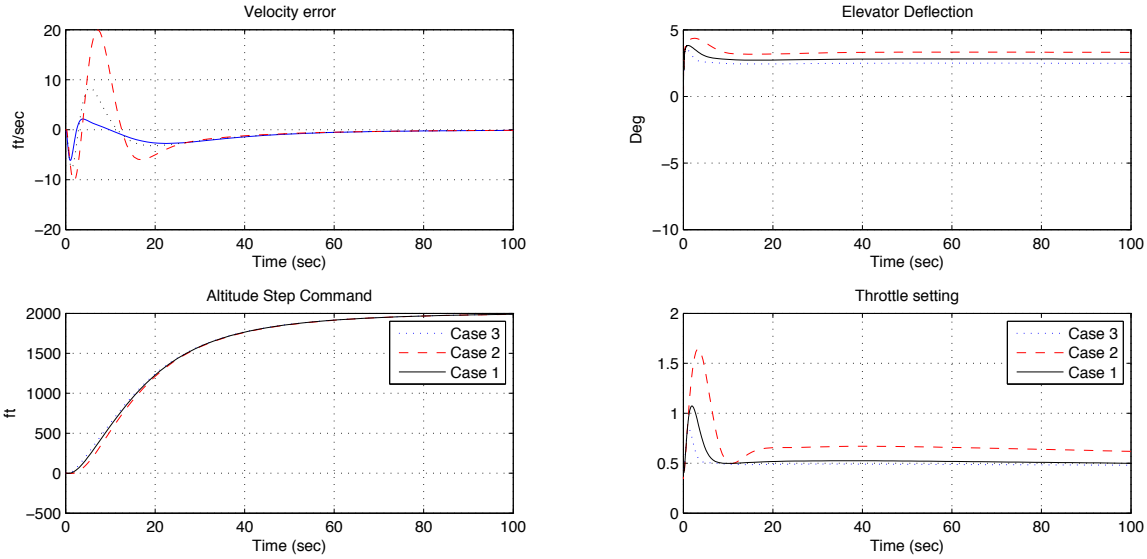
Another advantage of the minimax LQR controller is that this controller can be modified for a higher magnitude of uncertainties as long as they satisfy the following IQC condition.

$$\int_0^\infty (\parallel z_j(t) \parallel^2 - \parallel \zeta_j(t) \parallel^2) dt \geq -\chi^T(0) D_j \chi(0), \tag{44}$$

where $D_j > 0$ for each $j = 1, \cdots, 5$ is a given positive definite matrix.

Conclusions

A robust control scheme involving feedback linearization and minimax optimal control methods is proposed for an air-breathing hypersonic flight vehicle control problem. The proposed scheme utilizes the effectiveness of the widely used feedback linearization technique with special treatment of the non-vanishing nonlinear expressions, followed by a robust minimax optimal control (LQR) design method. A linearized uncertain system model for the AHFV longitudinal dynamics is derived and presented in a form which can be used in robust control design approaches. In the derivation of this model, twenty four inertial and aerodynamic parameters are considered to be uncertain with given upper bounds on their magnitudes.



(a) (b)

Fig. 5: (a) Response to a step change in velocity, (b) Control inputs.

The proposed linearized uncertain model is effectively used in synthesis using a minimax LQR tracking controller for velocity and altitude tracking of the AHFV. Controller responses are demonstrated to be very good for the given uncertainty bounds. The results of this simulation show that the proposed controller is robust for the uncertain system model and the states and inputs of the AHFV system remain within the practical limits during tracking. The presence of integrals of error states in the developed controller may cause integral windup problems, but such phenomena can be dealt with a variety of ways. In our work, this problem is alleviated by the introduction of a first order pre-filter.

Acknowledgments

This research was supported by the Australian Research Council.

References

[1] D. O. Sigthorsson and A. Serrani, "Development of linear parameter-varying models of hypersonic air-breathing vehicles," in *in Proc. AIAA Guidance, Navigation and Control Conference*, ser. AIAA-2009-6282, 2009,

[2] H. Xu, M. Mirmirani, and P. A. Ioannou, "Adaptive sliding mode control design for a hypersonic flight vehicle," *Journal Of Guidance Control, And Dynamics*, vol. 27, no. 5, pp. 829–838, September-October 2004.

[3] Q. Wang and R. F. Stengel, "Robust nonlinear control of a hypersonic aircraft," *Journal of Guidance Control, and Dynamics*, vol. 23, no. 4, pp. 577–585, July-August 2000,

[4] J. T. Parker, A. Serrani, S. Yurkovich, M. A. Bolender, and D. B. Doman, "Control-oriented modeling of an air-breathing hypersonic vehicle," *Journal of Guidance, Control and Dynamics*, vol. 30, no. 3, pp. 856–869, May-June. 2007,

[5] L. Fiorentini, A. Serrani, M. A. Bolender, and D. B. Doman, "Nonlinear control of non-minimum phase hypersonic vehicles models," in *American Control Conference*, St Loius USA, June 2009, pp. 3160–3165,

[6] O. Rehman, B. Fidan, and I. R. Petersen, "Robust minimax optimal control of nonlinear uncertain systems using feedback linearization with application to hypersonic flight vehicles," in *Proc. 48th IEEE Conference on Decision and Control*, Shanghai, China, 2009, pp. 720–726.

[7] ——, "Minimax LQR control design for a hypersonic flight vehicle," in *Proc. 16th AIAA/DLR/DGLR International Space Planes and Hypersonic Systems and Technologies Conference*, ser. AIAA-2009-7291, Bremen, Germany, Oct 2009.
[8] ——, "Uncertainty modeling for robust multivariable control synthesis of hypersonic flight vehicles," in *Proc. 16th AIAA/DLR/DGLR International Space Planes and Hypersonic Systems and Technologies Conference*, ser. AIAA-2009-7288, Bremen, Germany, Oct 2009.
[9] L. Fiorentini, A. Serrani, M. A. Bolender, and D. B. Doman, "Nonlinear robust adaptive control of flexible air-breathing hypersonic vehicles," *Journal of Guidance, Control and Dynamics*, vol. 32, no. 2, pp. 401–416, March–April. 2009,
[10] ——, "Robust nonlinear sequential loop closure control design for an air-breathing hypersonic vehicles," in *Proc. American Control Conf.*, Westin Seatle Hotel, Seatle, Washington, USA, June 2008, pp. 3458–3463,
[11] A. V. Savkin and I. R. Petersen, "Minimax optimal control of a uncertain systems with structured unertainity," *International Journal of Robust and Nonlinear Control*, vol. 5, no. 2, pp. 119–137, September-October 1995.
[12] A. Isidori, *Nonlinear Control Systems, 3rd Ed.* London: Springer, 1995.

HIFiRE Re-entry Observation Using an Image-Based Visual Servoing System

Razmi Khan [1], David Buttsworth [2], Ben Upcroft [1]

[1] *The Centre of Hypersonics,*
The University of Queensland, Brisbane, 4072, QLD, Australia
[2] *Faculty of Engineering and Surveying*
University of Southern Queensland, Toowoomba, 4350, QLD, Australia

Summary: This paper presents an image based visual servoing system that is intended to be used for tracking and obtaining scientific observations of the HIFiRE vehicles. The primary aim of this tracking platform is to acquire and track the thermal signature emitted from the surface of the vehicle during the re-entry phase of the mission using an infra-red camera.

The implemented visual servoing scheme uses a classical image based approach to identify and track the target using visual kinematic control. The paper utilizes simulation and experimental results to show the tracking performance of the system using visual feedback. Discussions on current implementation and control techniques to further improve the performance of the system are also explored.

Keywords: HIFiRE, Visual Servoing, Machine Vision, Visual Feedback

Introduction

HIFiRE is the Hypersonic International Flight Research Experiment - a joint venture between the Australian Defence, Science and Technology Organization and the US air force. The intent of the research is to ultimately achieve sustained hypersonic flight through a series of experiments which involve rocket-launching scramjet engines in a sequence of progressively challenging trajectories. There are considerable engineering challenges remaining in the area of scramjet-sustained hypersonic flight, particularly relating to thermal loads and supersonic combustion. The remaining flights within the HIFiRE program provide an excellent opportunity to glean additional data in these areas through ground-based observations. For example, emission spectroscopy through remote observation of the scramjet exhaust plumes could provide an additional diagnostic tool for scramjet combustion efficiency analysis. Furthermore, remote thermal IR imaging of the vehicle would also be possible because of the high surface temperatures which are achieved during flight.

Efforts to monitor the surface temperatures in hypersonic flight have previously been demonstrated on the HYTHIRM project [11-13]. This project successfully acquired high resolution thermal radiation images from the heat shield for several NASA Space Shuttle re-entries. The optical equipment mounted on an airborne platform used a hybrid tracking approach. Initial acquisition was done using computer aided techniques that would initially acquire the target within the field of view of the cameras. Once acquired in the FOV of the tracking camera, a mirror which reflected the radiation emitted from the heat shield surface was steered manually to keep track of the target centred in the image plane. Such position based systems typically require accurate equipment calibrations with the system being prone to target acquisition inaccuracies due to calibration errors and target position offsets. Manual

operation can be challenging especially when the target is at a large distance and high resolution imaging is required. Under these conditions, erroneous manual pointing of the camera equipment can arise and may cause failure to acquire the necessary data. To address this problem, an automated image based tracking approach is proposed to robustly obtain the emitted thermal radiation for the atmospheric re-entry of the HIFiRE test vehicle.

The proposed technique to track the re-entry involves implementation of classical image based visual feedback control based on experiments carried out by Corke et al. [8]. The approach uses a technique that would detect a manually chosen target between frames of a video sequence and track using a 2 axis pan and tilt robot as it moves across the field of view of the camera. Previously Hutchinson et al. [7] and Papanikolopoulos et al. [4] provided the fundamentals of visual servoing and discussed the control issues that affect stability and performance.

This paper discusses the vision and control strategies that are being developed for tracking the HIFiRE test vehicle on its hypersonic re-entry phase of the mission. System modelling will also be introduced to aid in selecting appropriate gains that would achieve a desired response to the system. Provisions for the further improvement of tracking performance of the system using feed forward control strategies will also be discussed. Note that the proposed thermal/IR camera in Figure 1 is for demonstration purposes only. Integration of suitable IR cameras with the proposed system is yet to be addressed.

The layout of the paper includes an outline of the equipment setup proposed and formation of an image Jacobian discussed in the following section. In section 'Visual Feedback Control' we model the feedback system and discuss the analysis and system timing. Following this we show the system tracking performance results and a comparison between the simulation and the actual system being developed. Finally, we conclude with a brief discussion on further improving the tracking performance of the system and implementation for the upcoming HIFiRE re-entry.

Visual Servoing Architecture

In this section we introduce the prototype equipment and setup. We also present the image Jacobian and the simplifications employed to analyze the control.

Equipment Setup

The prototype robot is a Directed Perception PTU-47-17 high speed pan and tilt unit. This unit hosts a 'Flea 2' black and white Point Grey camera capable of running at 30Hz. Figure 1 shows the mounting of this camera on the right and the proposed thermal camera on the left of the robot payload bracket. Initial indoor and outdoor target tracking tests were conducted using a 6mm focal length lens on the tracking camera. Note that the actual cameras and lenses that will be used for tracking and signal acquisition from HIFiRE have not been finalized. The equipment in the present work is adopted for the purpose of system analysis and demonstration. A typical desktop CPU interfaces with the camera and robot using an IEEE 1394b standard interface and serial respectively. All vision processing and robot movement tasks were handled by a C program written in Microsoft Visual Studios in a Windows based environment.

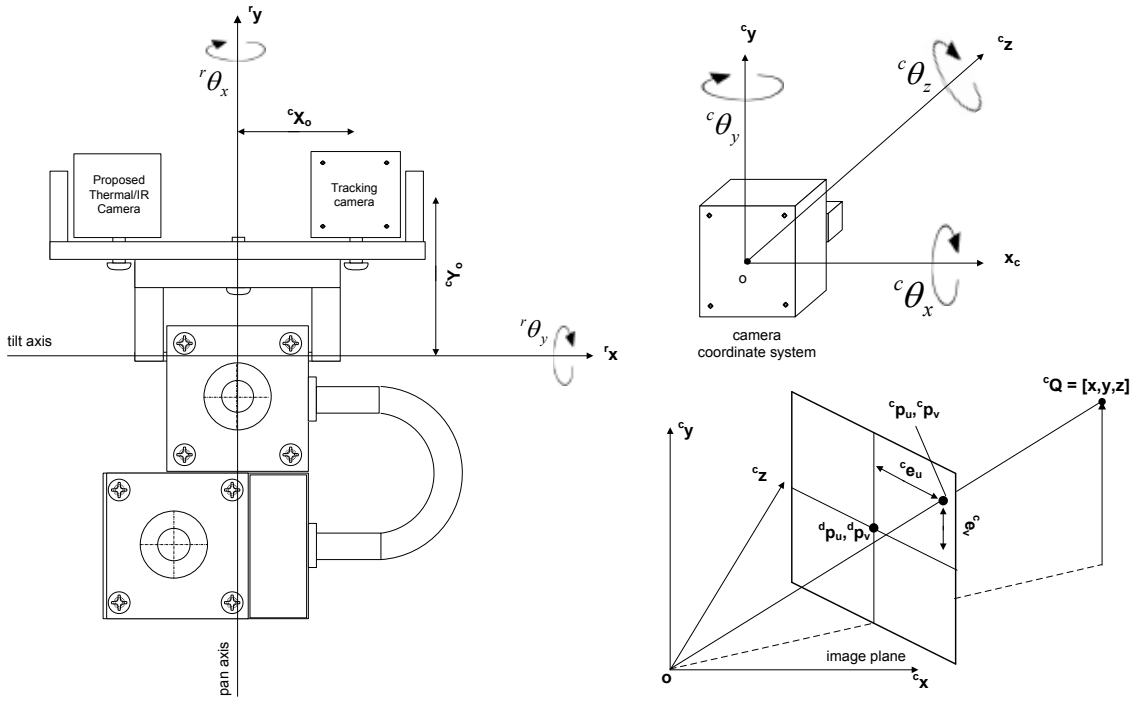


Fig. 1. A prototype of the robot (left) is shown with a 2-DOF rotational movement with proposed camera mountings. Figures on the right top and bottom demonstrate the camera coordinate system in terms of which all modelling will be shown and a typical target position on the image plane with an error condition respectively

Image Jacobian

For an end effector mounted camera setup, tracking a target requires a known relationship between the change in target position on the image plane to the change in robot pose. Using a pin hole camera model and a perspective geometric relationship between the target and the camera, an image Jacobian can be derived. Corke et al. [8] and Haralick et al. [5] show a detailed derivation of this transformation. The resultant image Jacobian for a point ${}^{c}\boldsymbol{O} = [x_c, y_c, z_c]$ expressed in the camera coordinates can be written as,

$$\dot{\boldsymbol{p}} = \boldsymbol{J} * \dot{\boldsymbol{r}}$$

which can be expressed as a matrix,

$$\begin{bmatrix} \dot{p}_u \\ \dot{p}_v \end{bmatrix} = \begin{bmatrix} \frac{f}{z} & 0 & \frac{-p_u}{z} & \frac{-p_u p_v}{f} & \frac{f^2 + p_u^2}{f} & -p_v \\ 0 & \frac{f}{z} & \frac{-p_v}{z} & \frac{-f^2 - p_v^2}{f} & \frac{p_u p_v}{f} & p_u \end{bmatrix} \begin{bmatrix} T_x \\ T_y \\ T_z \\ \omega_x \\ \omega_y \\ \omega_z \end{bmatrix}$$

where $\dot{\boldsymbol{p}}$ is the velocity of the image feature with p_u and p_v the target coordinates on the image plane. T_x, T_y, T_z and $\omega_x, \omega_y, \omega_z$ are translational and rotational velocities of the end effector respectively. f is the focal length of the camera lens in pixels.

The Jacobian, $\boldsymbol{J}$ can be simplified into two equations for a 2 DOF robot rotational motion only. For pure rotational motion in the pitch and yaw axis the Jacobian matrix can be reduced to the following equations,

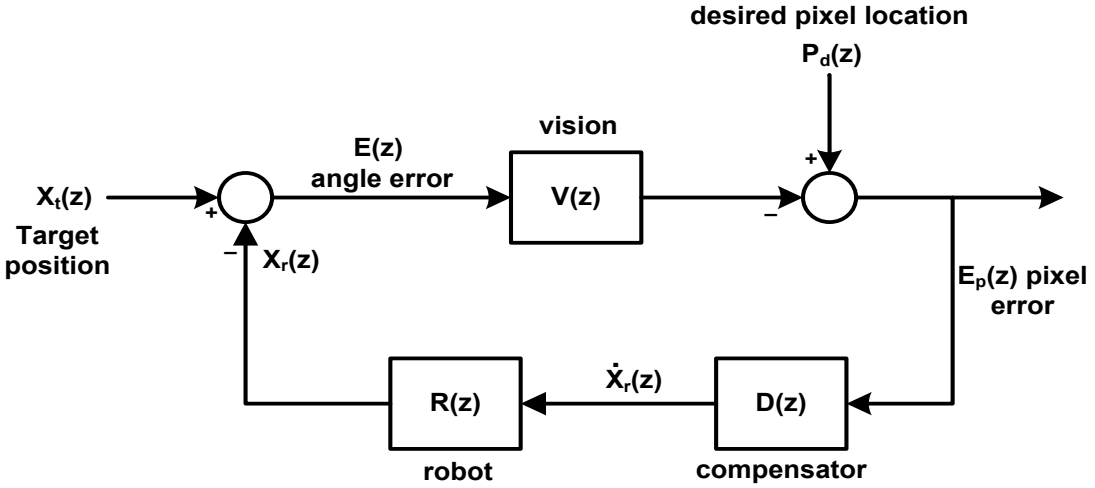


Fig. 2. Block diagram of the feedback control system. Note that the above figure shows single axis control only

$$\dot{p}_u = \frac{f^2 + p_u^2}{f}\omega_y \text{ and } \dot{p}_v = \frac{f^2 + p_v^2}{f}\omega_x$$

In forming the equations above, a few approximations need to be addressed. The off axis mounting of the tracking camera leads to the introduction of the two offset terms ${}^cX_o$ and ${}^cY_o$ (as shown in Figure 1) when transforming the robot coordinate system to the camera coordinates. The yawing and pitching motions about the y and x axes introduces robot translation velocities. Assuming that the distance of the object being tracked is ${}^c\boldsymbol{O} = [x_c, y_c, \infty]$ and the offset terms ${}^cX_o$, ${}^cY_o \approx 0$, the translational motion induced can be approximated as zero.

Vision algorithm

Visual servoing is typically classified into two phases. The first task requires the target to be identified between frames of a video sequence. The identified target (in this case a point object) is then centered within the frame by moving the end effector mounted camera. High resolution imagery of the re-entry vehicle is possible using a large focal length lens on the thermal cameras. Depending on the selected observation location with respect to the re-entering vehicle, the apparent size of the target will increase as the object moves closer to the imaging equipment. This may require the use of target identifying vision algorithms such as CAMSHIFT that take into consideration the increasing size of the object on the image plane. However, for simplicity all analysis and results discussed in this paper use targets as point objects and implements the Lucas Kanade optical flow pyramidal approach to acquire the target in subsequent frames [1].

Visual Feedback Control

In this section we present the visual feedback control scheme that will be used throughout our analysis and experimental sections. A high level architecture is shown in Figure 2. To simplify the analysis, the system will be assumed as a single axis system. Target movement is assumed only about the y axis that corresponds to the horizontal line on the 2D image plane. To

validate this assumption the 2-DOF prototype robot uses identical pitch and yaw axis stepper motors. Due to the insignificant weight and offset distances of the cameras from the centre of each of the rotational axis, the moment of inertia I_c can be approximated as zero.

System modelling

The prototype system is a single rate system that runs at the video update rate of 30Hz. The robot controller rate is programmable and is to set the same frequency. As shown by Corke et al. [6] a single rate discrete time system can be modelled primarily in terms of delay. Each module of the system is therefore represented as a function in discrete *z* domain notation.

The primary objective of the system is to detect and fixate as the target moves. Therefore, the output of the model shown in Figure 2 describes the image plane pixel error $E_p(z)$ for a given target motion $X_t(z)$. The desired pixel location $P_d(z)$ is always at the centre of the image plane with coordinates ${}^{d}p_u$, ${}^{d}p_v$. The discrepancies between the robot position and the target position in the image plane yields the error signal $E(z)$, in radians.

The compensator on the feedback path applies a fixed proportional gain to the image error velocity output $\dot{X}_r(z)$. The vision system, $V(z)$ is modelled as a single sample period delay, K_{lens}/z. The delay can vary between one to three frames depending on the number of cameras on the bus and the resolution of the cameras. K_{lens} is an approximation of the lens gain for a particular focal length set on the camera. For small angular target changes per sample period this gain can be approximated as the focal length of the lens in pixels/rad. The focal length was measured experimentally using the Camera Calibration Toolbox in Simulink MATLAB.

The compensator here is a purely proportional feedback gain. Together with the image Jacobian, the feedback provides the robot with a velocity command. The proportional gain K_p was chosen using analytical and experimental techniques to provide a critically damped response. These gains are analytically derived and discussed in the analysis section.

The robot $R(z)$ is composed of a combination of three parts, a robot dynamics $R_d(z)$ described in the next section, an integrator $z/z\text{-}1$, and a single frame delay $1/z$. The robot starts to change pose as soon as a command reaches the controller via serial communication and therefore moves in real time. The delay is introduced since the change in robot position is only sensed by the system at the next shutter event.

Robot dynamics

The robot acceleration and velocity information is provided by the manufacturer and is programmable to user defined settings.

The velocity of the robot $\dot{X}_r(z)$ is divided into two segments: velocities above and below base velocity V_b as shown in Figure 3. For velocities above the base speed the robot takes time to accelerate to the requested velocity and decelerate before reaching the requested pose.

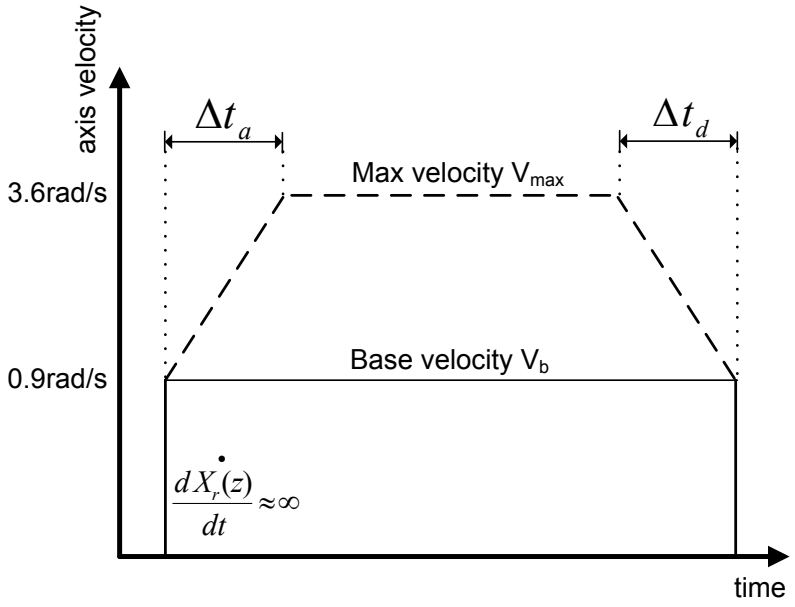


Fig. 3. Approximate velocity-time profile of the robot for with negligible load characteristics and moment of Inertia, $I_c \approx 0$

To simplify the analysis the current development of the prototype robot is only operated at base speeds resulting in an effective instantaneous acceleration and deceleration. This simplified dynamic behaviour of the robot can be modelled as a saturation function, *f(v)* where *v* is the input velocity command. The function *f(v)* is represented graphically in Figure 4 and is defined as,

$$f(v) = \begin{Bmatrix} v & if & -0.03 < v < 0.03 \\ 0.03 & if & v \geq 0.03 \\ -0.03 & if & v \leq -0.03 \end{Bmatrix}$$

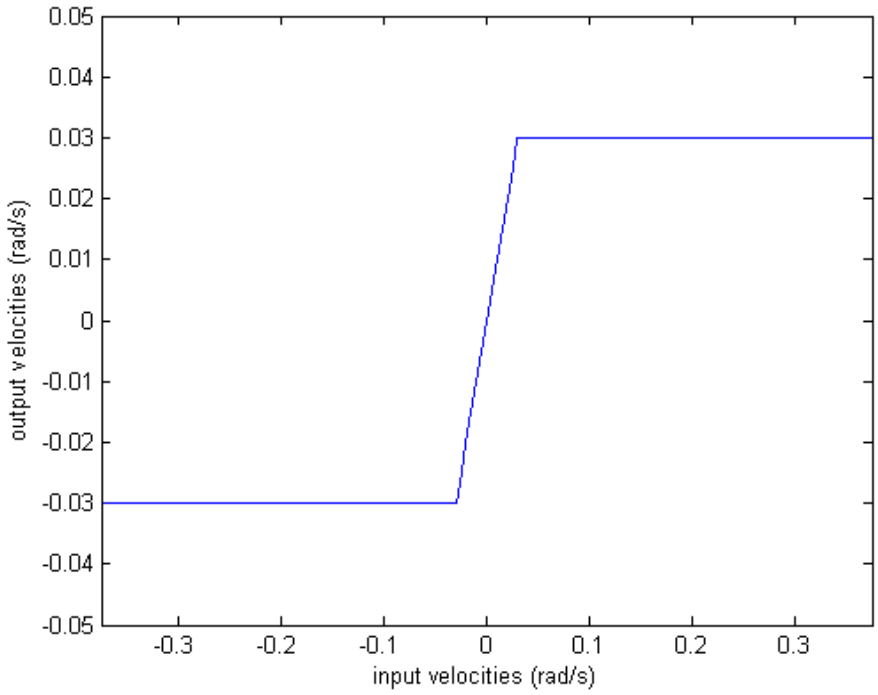


Fig. 4. Simplified robot dynamic model

Analysis of system modelling

The model developed in the previous section can be written as a closed loop transfer function that describes the image plane pixel error $E_p(z)$ for a given target motion $X_t(z)$,

$$\frac{E_p(z)}{X_t(z)} = \frac{V(z)}{1+V(z)R(z)D(z)}$$
$$= \frac{K_{lens}(z-1)}{z^2 - z + K_p / K_{lens}}$$

To analyse the stability of the closed loop system a root locus is drawn as shown in Figure 6 with the feedback proportional gain K_p of 1.

Note that the robot dynamics term is not included in the above transfer function. Due to the non-linearity of the function approximation, $f(v)$, the assumption of linearity in calculating the transfer function is invalidated. It is found that using a proportional controller only the robot dynamics will have no effect on the stability of the system. The term will only increase the amplitude of the oscillations shown in Figure 5 (dashed) while remaining marginally stable for a gain of 1. Therefore, the simplified dynamic model of the robot shown in Figure 5 will be excluded from the stability calculations.

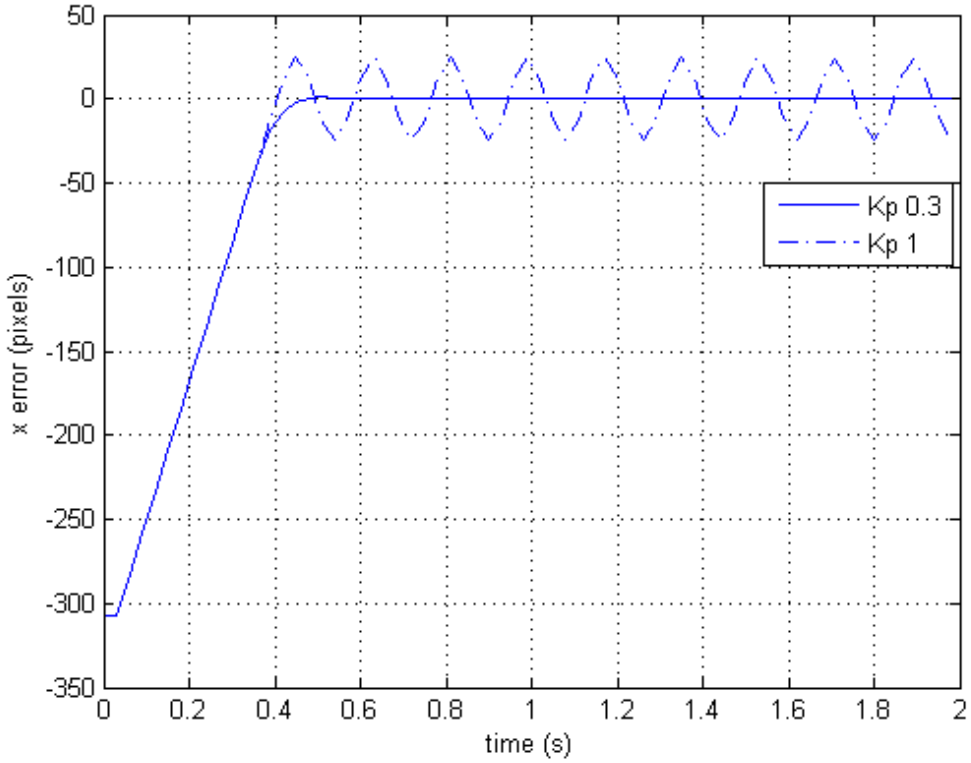


Fig.5. Simulated model response of Pixel Error E(z) to a 0.375 radian step input. The figure shows the critically damped and oscillatory response of the system for a proportional gain K_p 0.3 (solid) and K_p 1(dashed) respectively

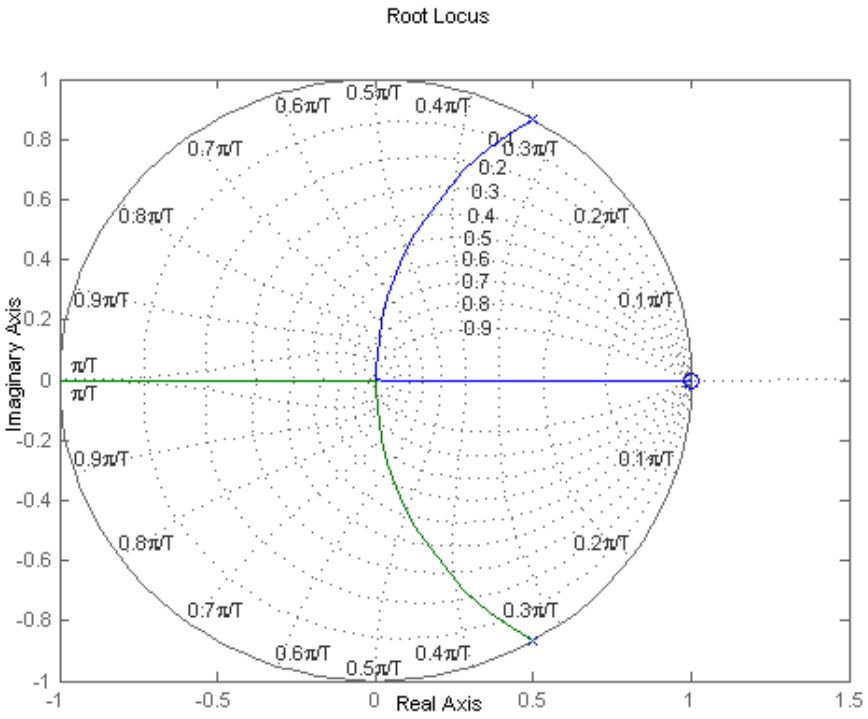


Fig. 6. Root locus plot for the closed loop visual servoing system with a feedback proportional gain K_p of 1.

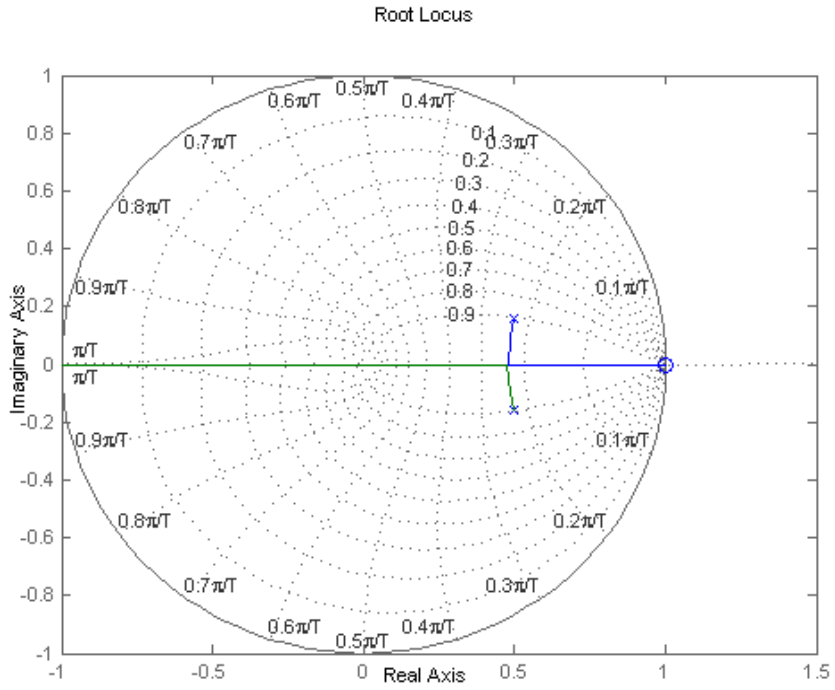


Fig. 7. Root locus plot for the closed loop visual servoing system with a feedback proportional gain K_p of 0.3. Note that the stability of the system increase at the cost of tracking performance

It is apparent from the root locus plot in Figure 6, that with a proportional gain K_p of 1 the complex poles of the closed loop system are on the unit circle. This results in a system that oscillates about a desired target location by the change in angle made by the robot in one sample period. The marginal stability in a closed loop visual servoing system is caused primarily by the latency in vision and is well documented by Corke *et al.* [8]. The following section describes the timing relationship of the system and discusses the reason for latency in this particular setup.

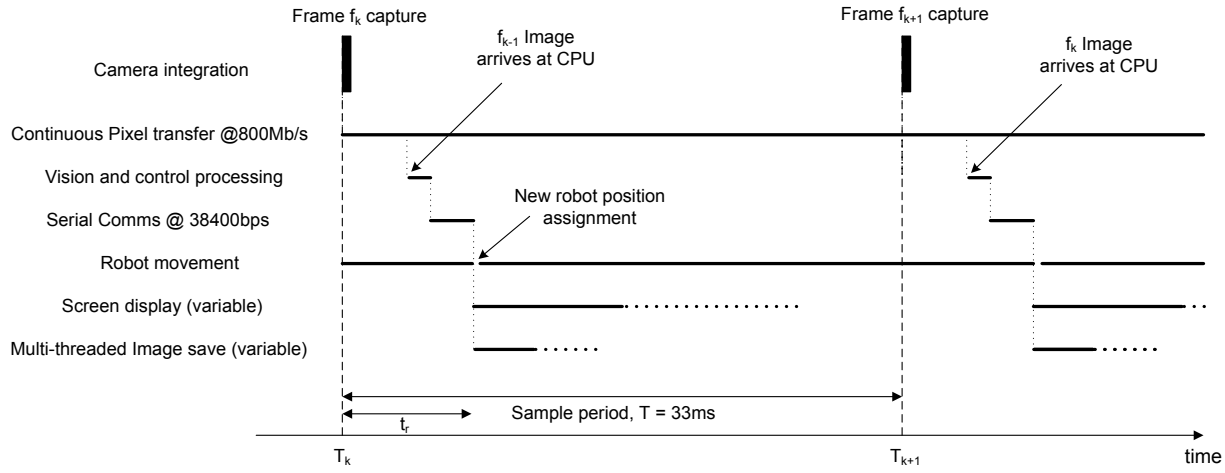


Fig. 8. System timing diagram for a large initial target movement where the robot axes cannot achieve the required velocities before the next shutter event. Note that the timing of each task is approximate and only provides a general idea about the times required for various tasks.

This instability is compensated at the expense of system performance. To obtain critical damping for a step change in the target angle, the gain K_p is set to 0.3. Figure 7 shows the change in the location of the poles with the difference between the two responses shown in Figure 5.

This configuration can achieve zero pixel error in approximately 420ms for a maximum step input, i.e. stationary target with offset from the centre of the image by 320 pixels when running on 640x480 pixel resolution.

System Timing

The behaviour and stability of this visual servoing system depends on the timing and completion of various tasks. Figure 8 shows the approximate duty cycle of tasks over a sample period *T*. It is clear that the primary cause of latency in this system is the pixel transfer which occurs between the camera and the host computer. The image processing algorithm and communication between the software and the robot takes approximately 4ms and is relatively small when compared to the delay caused by the transfer of pixels. Image f_k only arrives at the CPU after the f_{k+1} camera shutter event has elapsed, i.e. the image that is processed to determine robot angles uses target information out dated by a single frame. The latency was found to increase up to three sample periods when the number of cameras on the same bus or single camera resolution was increased[1].

For a high velocity target, the robot may not achieve sufficient rotation to keep the target centred in the frame for a given sample period. In this case the pose of the robot may still be changing while the subsequent frame is being captured as shown in Figure 8. This causes a short period of time, t_r where the robot rotation is not accounted for until the next frame and results in a slight overshoot of the robot position. The angle overshoot can be approximated as a rotational movement made by that axis in time t_r before the new robot position angles become available.

[1]The three sample period delay also includes a long shutter opening of the camera. All modeling and experiments are conducted using an ideal sample.

The vision algorithm and the control are only activated by the use of a single mouse click event. The click serves as a manual target detection task which activates the vision and control processing on each of the following frames. The initial click may occur stochastically between a shutter event[1] and cause variable robot axis movement for the initial 'tracking activated' sample period. Following this initial frame the robot axes movements are allowed to change positions for a complete 33ms allowing fixed times for the robot to move. This validates the robot movement of 0.03rad/33ms assumed earlier to form a simplified robot dynamic model.

Experiments

This visual servoing system is tested using various target trajectories to illustrate the tracking performance of the system. Comparison between the actual system and simulations are also discussed to illustrate the accuracy of the simulation model. The assumption made during the design stages especially in approximating the moment of inertia to zero is validated by the similarities in the response of the simulation when compared to the actual system.

Figure 9 (top) shows the tracking error performance for a sinusoidal target trajectory for the pan axis. Pendulum motion of the target was set at an angular velocity of approximately 2.5 rad/s, with peak amplitude of 0.25 radians. The system fixates on the target with a maximum

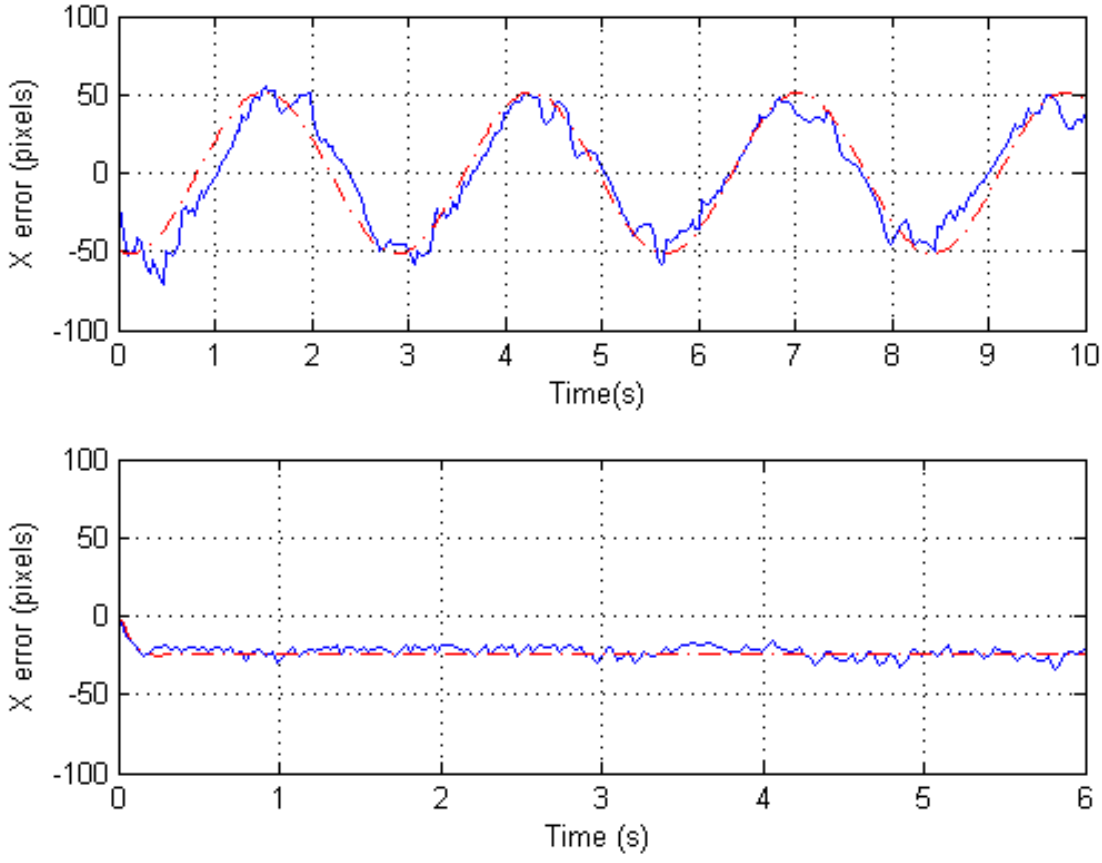


Fig. 9. Response of the system(solid) and simulation (dashed) to different target trajectories. The plots show the x axis pixel error response of the system for a constant velocity (bottom) and a sinusoidal (top) target trajectory at 640x480 pixel resolution. The target movement is approximately 0.25 radians in amplitude and with a frequency of 2rad/s for sinusoidal and 0.5rad/s continuous for constant velocity target trajectory.

[1]Shutter event is referred to the opening and closing of the camera shutter

error of 50 pixels when running at a 640x480 pixel resolution. Note that when the pixel resolution is increased the error increases due to higher resolution and also due to the increase in the number of pixels to be transferred from the camera to the CPU causing increased latency in the vision system.

Experiments were also conducted using a constant velocity motion of the target set at 0.5rad/s. A tracking steady state error of about 23 pixels is obtained when a single camera is operated at 640x480 pixel resolution.

Conclusion

This paper discusses the implementation of a visual feedback control scheme that is being developed for tracking and obtaining thermal radiation data of HIFiRE vehicles. The initial development of the feedback control scheme shows the advantages of using visual servoing over passive target tracking methods and manual crew operations. The scheme uses much simpler equipment and system design to perform the tracking task. The most significant factor affecting the performance of the system is the delay caused by the latency in the vision system.

To reduce the steady state error in the constant velocity and peak to peak error in the sinusoidal motion target tracking can be reduced significantly using feedforward control architecture. This technique uses estimation to predict current target location based on the previous movements and overcome the lag caused by latency of the vision system discussed earlier. Such feedforward control design can be implemented using existing equipment in the final ground based system used for the observation. It is anticipated that in future, the observation equipment would be installed on an unmanned airborne platform that could fly closer to the vehicle trajectory and obtain high resolution thermal imagery and emission spectroscopy autonomously.

References

1. B.D.Lucas and T.Kanade, “An image registration technique with an application to stereo vision”, in *Proceedings International Joint Conference on Artificial Intelligence*, 1981, pp.674-679
2. Roger Y.Tsai and Reimar K.Lenz, “A New Technique for Fully Autonomous and Efficient 3D Robotics Hand/Eye Calibration”, *IEEE transaction on Robotics and Automation*, vol 5, np.3, June 1989
3. J.T.Feddema, C.S.G Lee, and O.R.Mitchell, “Weighted Selection of Image Features for Resolved Rate Visual Feedback Control”, *IEEE transaction in Robotics and Automation*, vol 7, pp.31-47 Feb 1991
4. N.P.Papanikolopoulos, P.K Khosla, and T.Kanade, “Visual tracking of a moving target by a camera mounted on a robot: A combination of vision and control”, *IEEE transaction in Robotics and Automation*, vol 9, no.1, pp.14-35 1993
5. R.M.Haralick and L.G. Shapiro, “Computer and Robot Vision”, Reading, MA: Addison-Wesley, 1993
6. Peter I.Corke, “High Performance Visual Closed Loop Robot Control”, *PhD dissertation*, University of Melbourne, Department of Mechanical and Manufacturing Engineering, July 1994
7. Seth Hutchinson, Gregory D.Hager and, Peter I.Corke, “A Tutorial on Visual Servo

Control", *IEEE transactions on Robotics and Automation*, vol. 12 no. 5, October 1996
8. Peter I.Corke and Malcolm C.Good 1996, "Dynamic Effects in Visual Servoing Closed-Loop System", *IEEE transaction in Robotics and Automation*, vol 12, no.5, p671-683 October 1996
9. Alessandro De Luca, Giuseppe Oriolo and Paolo Robuffo Giordano, "Feature Depth Observation for Image-based Visual Servoing: Theory Experiments", *The International Journal of Robotics Research* 2008;27;1093
10. Odile Bourquardez, Robert Mahony, Nicolas Guenard, Franc¸ois Chaumette, Tarek Hamel, and Laurent Eck, "Image based Visual Servo Control of the Translation Kinematics of a Quadrotor Aerial Vehicle", *IEEE transaction in Robotics and Automation*, vol 25, no.3, June 2009
11. Thomas J.Horvath, Deborah M.Tomek, Karen T.Berger, Scott C.Splinter, Joseph N.Zalameda, Paul W.Krasa, Steve Tack, Richard J.Schwartz, DavidM.Gibson, and Alan Tietjen, "The HYTHIRM Project: Flight Thermography of the Space Shuttle During Hypersonic Re-entry", AIAA 2010-241, *48th AIAA Aerospace Science Meeting* Jan 4-7, Orlando, Florida, 2010
12. Steve Tack, Deborah M.Tomek, Thomas J.Horvath, Harry A.Verstynen, Edward J.Shea, "Cast Glance Near Infrared Imaging Observation of the Space Shuttle During Hypersonic Re-entry", 2010-243, *48th AIAA Aerospace Science Meeting*, Jan 4-7, Orlando, Florida, 2010
13. Joseph N.Zalameda, Alan B.Tietjen, Thomas J. Horvath, Deborah M.Tomek, David M.Gibbs, Jeff C.Taylor, Steve Tack, Brett C.Bush, David Mercer, Edward J.Shea, "Application of a Near Infrared Imaging System for Thermographic Imaging of the Space Shuttle during Hypersonic Re-entry", 2010-245, *48th AIAA Aerospace Science Meeting* Jan 4-7, Orlando, Florida, 2010

Processes Forming Volcanic Topography at Atla Regio, Venus

Elyse Schinella[1], Craig O'Neill[1], Juan Carlos Afonso[1]

Department of Earth & Planetary Science, GEMOC ARC National Key Centre, Macquarie University, Sydney, NSW, 2109

Summary: Venus and Earth share a similar size, mass and density, however Venus' high surface temperatures, pressures, dehydrated crust and absence of subduction zones, makes the surface unique. One area which shares similar volcanic landforms to the Earths' is the equatorial highland Alta Regio, which extends from 170°E-220°E 30°N-10°S. The similarity of these volcanic features to Earth based examples, as well as an understanding of the processes which produced them, can help to constrain the deformational mechanisms that may have operated at Atla Regio. This study will aim to compare volcanic topography at Atla Regio to topography observed within the Hawaiian and Cape Verde Archipelagos, Earth to determine if similar deformational processes occurred within Atla Regio. On Earth, three main processes act to produce volcanic topography and include: 1) uplift associated with a mantle plume impinging on the underside of the lithosphere; 2) rifting and volcanism associated with the mantle plume and; 3) volcanic loading, where the extra volcanic mass is compensated by flexure of the lithosphere. An understanding of the processes that gave rise to topography at Atla Regio is fundamental to determine a suitable rheological profile within this area.

Keywords: Atla Regio, Venus, Hawaii, Cape Verde, mantle plume, volcanism, lithospheric flexure

Introduction

Venus consists of three main types of geological landforms, including highland areas, lowland volcanic plains and lowland plains regions, which encompass 8%, 27% and 65% of Venus' surface, respectively [1]. One highland located along the equator is Atla Regio, which is marked by three large volcanoes, Ozza Mons in the south-east, Maat Mons in the south-west and Sapas Mons in the north-west, as well as three rifts, Ganis Chasma, Dali Chasma and Parga Chasma that form a triple junction at Ozza Mons (Figure 1) [2]. Due to the presence of these rifts and volcanoes, in conjunction with this region's broad domal morphology, gravity high, extensive lava flows, high elevation (~3 km) and absence of compressional features, Atla Regio is considered to be a volcanic rise [2]. Atla Regio shares similar characteristics to Earths' hot-spots, where volcanism is associated with mantle plumes [1, 2]. On the basis of the localised topography and gravity observed at Ozza Mons and Sapas Mons, each of these areas may be sites of mantle upwellings, implying Atla Regio may have two mantle plumes acting to produce the topography and gravity anomalies within this area [2].

Volcanic topography on Earth can result when an upwelling mantle plume impinges onto the base of the lithosphere causing volcanism and uplifted topography forming a volcanic rise [3] (Figure 2). Generally these areas also have large dimensions > 1000 km, suggesting that the topography is not solely supported by lithospheric strength, but by a deeper source such as a mantle plume [4]. Two currently active, Earth hot-spots are Hawaii, situated along the Hawaiian-Emperor Seamount chain and Cape Verde, located west of Senegal, Africa [5, 6]. Currently Hawaii and the Cape Verde Islands are located above mantle plumes, which cause

uplifted topography and a volcanic rise [5, 6, 7, 8].

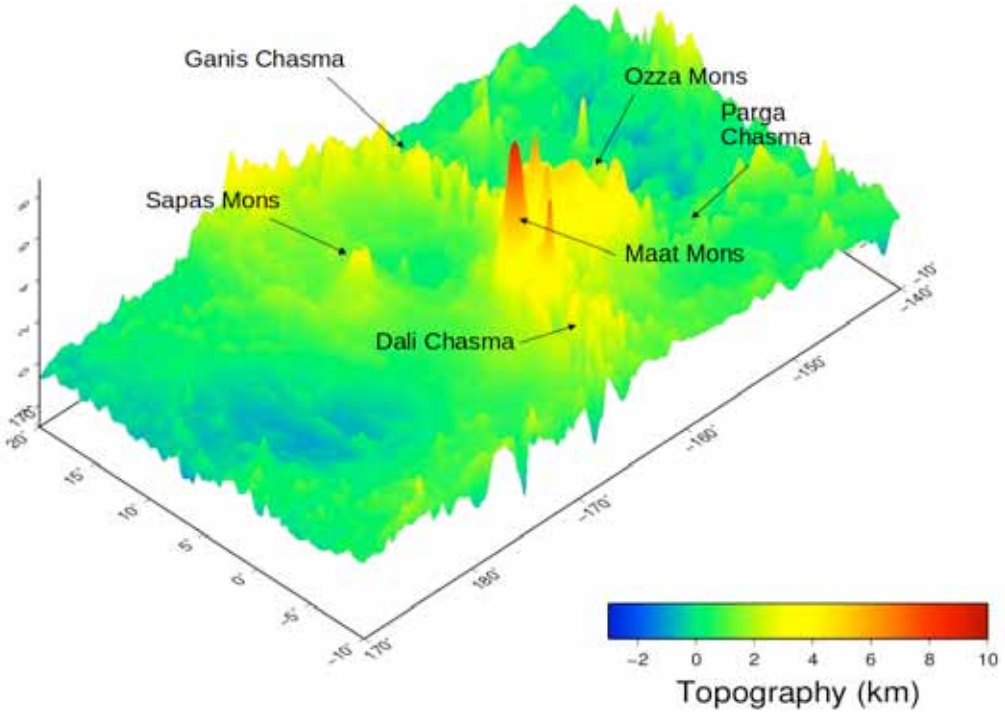


Figure 1 A 3D map showing the topography at Atla Regio. This area is dominated by three large volcanoes; Ozza and Maat Mons to the south and Sapas Mons to the north. A triple junction rift, comprising Ganis, Parga and Dalia Chasma meet at Ozza Mons.

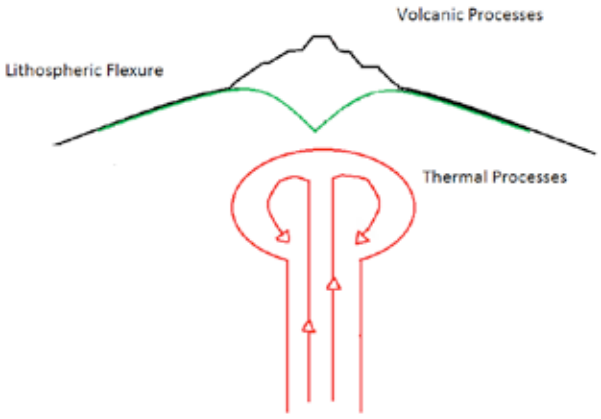


Figure 2 Schematic showing the three processes that can give rise to volcanic topography. Red line is an upwelling mantle plume, which can produce uplift, volcanism and the production of volcanic topography. The volcanoes produced can act as loads and cause the lithosphere to bend (green line).

Aside from uplift and volcanism, the presence of a mantle plume can be further constrained by an excess temperature, since mantle plumes may represent extra heat, which is superimposed on the background temperature associated with mantle convection [8, 9]. These temperatures rely on a stagnation distance (r_s), which is the distance that separates the upwelling plume material from the normal asthenosphere and a plume channel thickness (A) , which is the diameter of the asthenospheric conduit that connects the plume from the core-

mantle boundary to the base of the lithosphere [8, 9]. Sleep [8] calculated an average excess temperature of 230-300 °C for stagnation distances of 350-450 km and a plume channel thickness of 100 km for the Hawaiian swell.

The origin and structure of mantle plumes on Venus is highly unconstrained with some authors proposing that plume channels may not extend throughout the entire depth of the mantle [10]. However, if a plume channel exists below Atla Regio it should be thicker (~200±100 km), than those estimated for Hawaii (~100 km±16 km), due to Venus' hotter mantle [8, 11, 12]. Also coronae are possible surface expressions of mantle plumes, therefore their range of diameters 200-600 km could represent a range of possible stagnation distances [13].

The presence of a mantle plume can also be associated with a buoyancy flux, where a thermal mass is upwelling to the lithosphere per second, by a plume [8, 9, 12]. Since plume buoyancy is proportional to heat flow, the excess heat flow associated with this plume can also be calculated [8, 9, 12].

For the Earth, the calculation of buoyancy flux relies on the knowledge of the absolute velocity of a hot-spot, which is defined as the motion of the hot-spot relative to the plate it resides on [8, 9]. Therefore, lower heat fluxes associated with the mantle plume component of heat flow are on the order of 10-20 mW m^{-2} at slower-moving hot-spots, such as Cape Verde and Bermuda, compared to the higher mantle plume heat flux obtained at faster-moving hot-spots, such Hawaii [8, 9, 14].

Turcotte & Schubert [12] used a plate velocity of 90 mm yr^{-1} for the Hawaiian hot-spot with a cross-sectional area of 1.13 km^2, to calculate a buoyancy flux of 7.4 Mg s^{-1}, which represents a mantle plume heat flux of $3x10^{11}$ W or less than 1% of total global heat flow of the Earth at this location. Venus, however, has a stationary lithosphere and the closest Earth analog for these features may be for a slow-moving plate, since these areas may be closer to thermal equilibrium [14, 15]. Cape Verde is one example of a hot-spot located on a slow-moving plate [8, 14, 15]. The buoyancy flux, and in turn, the heat flow due to a mantle plume within this area is 1.6 Mg s^{-1} and 20 mW m^{-2} [8, 14]. Models by Smrekar & Parmentier [16] suggest that buoyancy fluxes, and in turn the extra heat fluxes from a mantle plume, which are most representative of Venus hot-spots, may not be large [14, 15].

Volcanoes associated with the impingement of a mantle plume can also initiate rifting and volcanism, contributing to the topography observed [17]. Hawaii is dominated by shield volcanoes, which are characterised by their broad shapes, low elevations and undulating slopes [17]. This morphology is also consistent with Venusian volcanoes, however, these volcanoes generally have larger basal diameters and flatter profiles in comparison to those found at Hawaii [18]. The larger basal diameters of Venusian volcanoes could be caused from an absence of plate motion, since a mantle plume would produce volcanism for longer time frames on a stationary lithosphere [1]. The lower elevations and shallower slopes of Venusian volcanoes could result from: 1) high temperatures within the crust and at the surface that may have acted to slow cooling of rising magma; 2) low lava viscosities; 3) highly effusive, large volumes of magma; 4) extensive lava flow formation from lava tubes; 5) slow cooling of lava flows due to high atmospheric temperatures [18, 19, 20].

Modelling of gravity anomalies over Alta Regio can be useful to determine a suitable density distribution and therefore, geological structure below the highland [13]. The highland topography associated with volcanism at Atla Regio should correspond to a high free-air gravity anomaly, high geoid anomaly and a low bouguer anomaly [2]. Free-air gravity

anomalies at short-wavelengths correspond well to local topography [12]. Bouguer anomalies, however, tend to exhibit structure associated with long-wavelength, isostatically compensated topography [12]. At sites of large-scale topography a negative bouguer anomaly is usually observed since the excess mass of the highland is isostatically compensated by a low-density crustal "root" or a low-density partial melt and magma chamber associated with volcanism [12]. The geoid anomaly, defined as the elevation difference between a reference equipotential surface (geoid) and the measured geoid, also reflects deeper density anomalies [12, 21, 22].

The presence of these volcanoes on the surface can act as loads which force the lithosphere downwards at the center of the load, whilst the surrounding area bulges upwards (Figure 3) [12]. By modelling this flexural response the elastic lithosphere thickness, which is the portion of the lithosphere that is rigid enough to sustain elastic stresses for long periods, can be determined [12]. This elastic lithosphere thickness can be used to determine a suitable temperature at the base of this layer, and in turn a thermal gradient and heat flux through an area [23].

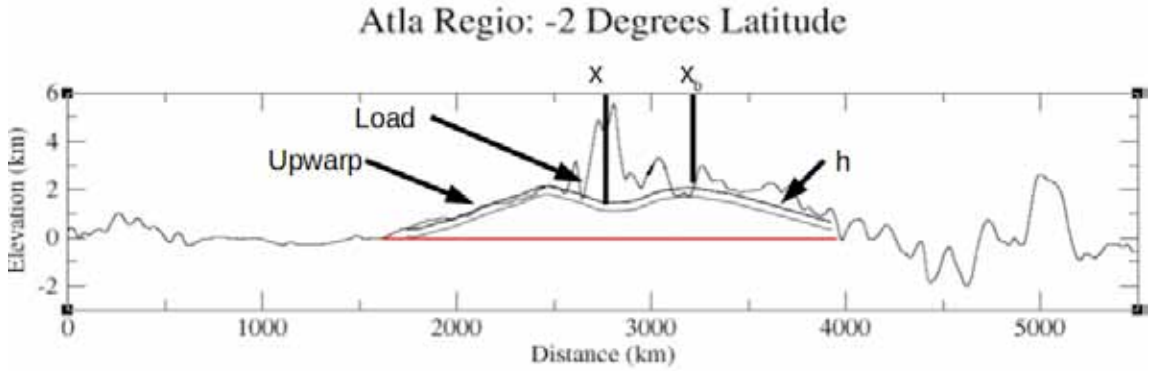


Figure 3 An east-west profile showing the upwarping of the lithosphere associated with a volcanic load at Atla Regio. The distance between the center of the load (x) and the maximum amplitude of the forebulge (x_b) can be used to determine elastic lithosphere thickness (h). The red line is the inferred line of constant depth.

Previous authors have determined a range of elastic lithosphere thicknesses for Atla Regio to reside between 20-52 km [14, 15, 24]. Even though these authors used spectral analysis of gravity and topography to determine elastic lithosphere thickness, their estimates are valuable for comparison purposes [14, 15, 24]. Phillips [15] used spectral analysis and Monte Carlo modelling to suggest a range of elastic lithosphere thicknesses between 40-50 km and a mean elastic lithosphere thickness of 45 km. Using this elastic lithosphere thickness as a guide, a thermal gradient of 7-10 K km^{-1} was found from moment-curvature relationships [15]. The elastic thickness estimate for Atla Regio was consistent with those obtained over the uplifted portion of the slow-moving, East African hot-spots (43-49 km), which experience a total heat flow (from mantle convection and a possible mantle plume) of 20-50 mW m^{-2} and a temperature gradient of 5-12.5 K km^{-1} [15, 25].

Phillips [15] inferred Atla Regio to share a similar thermal environment as the East African hot-spots and used an Earth-scaled heat flux of 74 mW m^{-2} as a guide to determine total heat flow within this area. Assuming a lower bound total heat flow of 80 mW m^{-2} for Atla Regio and a temperature gradient of 20 K km^{-1}, an elastic lithosphere thickness of 20 km was found [15]. However, this elastic lithosphere thickness is significantly lower than spectral analysis suggested, which led Phillips [15] to propose that Atla Regio should have a lower total heat

flow (i.e. < 80 mW m^{-2}) than what is ascertained from Earth scaling alone.
Phillips [15] also used the mean elastic lithosphere thickness and thermal gradient within Atla Regio to propose a 100 km thick thermal lithosphere, which comprises the crust and top-most portion of the upper-mantle, to occur below this area. A thin thermal lithosphere below Atla Regio, compared to the plains regions (~350 km) could suggest that the area experiences possible heating from the tail of a mantle plume [15].

Turcotte & Schubert [12] estimated the elastic lithosphere thickness below the Hawaiian Archipelago, by taking the distance from the center of the Island of Oahu to the surrounding arch. Using this distance (250 km), a flexural parameter and rigidity of 80 km and $2.4x10^{23}$ N m, respectively, an elastic lithosphere thickness of 34 km was found for this area [12]. The total heat flux at the Hawaiian swell was found from measurements to be ~52.9 ± 2.3 mW m^{-2} by Von Herzen *et al.* [26].

Spectral admittance studies suggest that the thickness of the elastic lithosphere may vary between 20-29 km below the Cape Verde Islands [6]. Best fit surface/subsurface loading models, however, suggest an elastic lithosphere thickness closer to 29 km [6]. Heat flow measurements taken across the Cape Verde Rise were found to increase over the middle of the swell up to 16±4 mW m^{-2} above the 45.5±3.4 mW m^{-2} heat flow of normal 125 Myr old crust [27].

Methodology

Parameters used to calculate the excess temperature associated with a mantle plume at Atla Regio, Oahu and Cape Verde are listed in Table 1.

Table 1 Parameters used to calculate excess temperature, buoyancy flux and heat flow.

	Parameter	Atla Regio (Venus)	Oahu (Earth)	Cape Verde (Earth)
Cross-sectional area (km^2)	W	3056*	1188*	1097*
Mantle density (kg m^{-3})	ρ_m	3330**	3330‡	3330‡
Surface density (kg m^{-3})	ρ_s	67□	1030†	1030†
Stagnation distance (km)	r_s	200-600□	350 – 450‡	390‡
Plume channel thickness (km)	A	200	100‡	100‡
Thermal expansion coefficient ($x10^{-5}$ K^{-1})	α	3.10#	3.00‡	3.00‡
Specific Heat ($x10^3$ J kg^{-1} K^{-1})‡	c_p	1.25	1.25	1.25

* Values measured from topographic profiles.
** Value from [28].
‡ Values from [7].
□ Density of Venus' atmosphere at the surface. Value from [29].

† Density of water. Value from [7].
□ Values from [13].
Value from [30].

These parameters, as well as Equation [1] can be used to estimate the excess temperature (T) due to a mantle plume:

$$T = (W(\rho_m - \rho_s))/(\pi\ r_s\ A\ \rho_m\ \alpha) \quad [1]$$

where W is the cross-sectional area, ρ_m is mantle density, ρ_s is surface density, which is considered to be atmosphere for Venus (67 kg m^{-3}) and water for Earth (1030 kg m^{-3}) and α is the coefficient of thermal expansion [8, 9].

The excess mass produced due to the presence of the plume can be represented by the buoyancy flux (B) delivered to the rise by the plume:

$$B = [(\rho_m - \rho_s)\ W\ v] \quad [2]$$

where W is the cross-sectional area and v is the velocity of the hot-spot relative to the lithosphere. This equation follows from Sleep [9] and the terminology in this equation is not standardised. Venus has a stationary lithosphere, therefore the closest Earth-based hot-spot analogue would be for a slow-moving plate [14]. The v values used for Venus are the lower limit plate velocity values of the slow-moving Cape Verde hot-spot (12 mm yr^{-1}) [27]. Hot-spot velocities relative to the lithosphere for the Hawaiian swell range from ~83 mm yr^{-1} to 96 mm yr^{-1} [4, 9, 31, 32]. A value of 90 mm yr^{-1} was used, similar to Turcotte & Schubert [12], since this value is closer to the average plate velocity for the values from [4, 9, 31, 32].

The heat flux (Q_H) of the mantle plume could then be found using:

$$Q_H = (c_p\ B\ ) / \alpha \quad [3]$$

The properties of volcanoes, which were calculated from topographic profiles over Atla Regio and the Hawaiian and Cape Verde swells, included basal diameter, height, average slope, root mean square (RMS) slope and volume. Topographic profiles across the edifice of the volcano were observed and the basal extents defined where the edifice joined the surrounding volcanic flanks. The heights, volumes, average slope and RMS slope of the volcano were then measured above the edifice base.

Modelling of the lithospheric structure below Atla Regio was completed using the finite-difference code GEO3Dmod [21]. GEO3Dmod is an interactive 3D forward modelling software, which calculates the thermal, pressure and density structure for a given model, as well as elevation, free-air gravity, bouguer gravity, geoid and heat flow [21]. By fitting these calculated values to observables, the models are more tightly constrained than by singularly fitting each observable [21, 22, 33]. GEO3Dmod assumes a conductive thermal lithosphere, comprising the crust and the conductive portion of the upper mantle [21, 22, 33]. See [21, 22, 33] for more details.

Lithospheric flexure calculations were also used to determine the thermal gradient and in turn heat flow through an area. The half-width of the depression (x) was located beneath the load and the distance (d) to the forebulge (x_b) calculated (Figure 3).

This distance (d) was then used to determine the flexural parameter (F) using:

$$F = d / \pi \quad [4]$$

Using the flexural parameter (F), the flexural rigidity (D) of the plate was found:

$$D = F^4 (\rho_m - \rho_s) g \quad [5]$$

where g is the acceleration due to gravity (9.81 m s^{-2} for the Earth and 8.87 m s^{-2} for Venus).

The flexural rigidity was then used to estimate the elastic lithospheric thickness (T_e) from:

$$T_e = [D \times 12 (1 - \nu^2) / E]^{1/3} \quad [6]$$

where v is Poisson's ratio (0.25) and E is Young's modulus (70 GPa), similar to the values used by [12].

By using a reference lithosphere and moment-curvature relationships, Phillips *et al.* [14] determined a suitable equation to determine a linear thermal gradient (dT/dZ) for elastic lithosphere thicknesses below certain Venusian features:

$$dT/dZ = 9.54 (T_e / 30)^{-0.817} \quad [7]$$

Using this temperature gradient (dT/dZ) and the thermal conductivity (k) at the base of the lithosphere, the total heat flux (q) of the area was determined by using the equation:

$$q = k \text{ -}dT/dZ \quad [8]$$

Estimates of the thermal conductivities at the base of the lithosphere were ~3.3 W m^{-1} K^{-1} [30] for Venus and 2.55 W m^{-1} K^{-1} for Hawaii [34, 35]. It should also be noted that the total heat flux calculated by Equation [8] includes both a heat flux from a possible upwelling mantle plume and the background heat flux associated with mantle convection [14].

Results

From Equation [1] the excess temperature associated with a mantle plume below Atla Regio was found to range from 85-769°C for stagnation distances between 200-600 km. Venus has been proposed to have a stationary lithosphere (plate velocity = 0 mm yr^{-1}). However, due to the dependency of buoyancy flux and heat flow equations on a plate velocity, a hot-spot located on a slow-moving plate was considered to be the closest Earth-based analogue [9, 14].

By using a lower-limit velocity (12 mm yr^{-1}) of the Cape Verde hot-spot, relative to the lithosphere [31], and Equation [2-3], a buoyancy flux of 3.79 Mg s^{-1} and a mantle plume heat flux of 1.53×10^{11} W at Atla Regio was obtained. Therefore, the the range of excess heat associated with a mantle plume on Venus would be less than 1.53×10^{11} W.

The excess temperature associated with Oahu, a fast-moving hot-spot, and was found to range from 193-249°C for stagnation distances between 350-450 km. Based on a plate velocity of 90 mm yr^{-1} [11], a buoyancy flux of 7.79 Mg s^{-1} and a mantle plume heat flow of 3.25×10^{11} W was found for Oahu. The excess temperature of Cape Verde, a slow-moving hot-spot, was found to be 206 K for a stagnation distance of 390 km. Using a plate velocity of 12 mm yr^{-1}, a buoyancy flux of 0.96 Mg s^{-1} and a heat flow of 4.0×10^{10} W was found.

Results for calculating the basal diameter, height, average slope and RMS slope for the three volcanoes within Atla Regio and comparing these to Hawaii and Pico de Fogo (within the Cape Verde Archipelago), can be observed in Table 2. The volcanoes located at Atla Regio have larger basal diameters (405-630 km), and except for Sapas Mons, larger volumes (30.5-

$42.3x10^4$ km^3), than the Island of Hawaii and Pico de Fogo, which have a basal diameters of 200 km and 63 km and volumes of $12.5x10^4$ km^3 and $1.46x10^4$ km^3, respectively.

All the volcanoes located within Atla Regio also have smaller heights (2.34, 3.85 km, 6.8km for Sapas, Ozza and Maat Mons, respectively) when compared to Hawaii (8.48 km). However, only Ozza Mons and Sapas Mons have smaller heights when comparing to Pico de Fogo (6.28 km). The RMS slope of each volcano was also calculated, however, since long slope-frequency distributions will bias the RMS slope to greater values, an average of the slopes measured for each volcanic edifice were also calculated [36, 37]. Shallower RMS slopes and average slopes were obtained for the volcanoes within Atla Regio (0.646°-2.19° and 0.537°-1.66°), compared to Hawaii (5.60° and 5.01°), and Pico de Fogo (13.55° and 12.06°). Since Pico de Fogo is a stratovolcano it has steeper slopes compared to the volcanoes within Hawaii and Atla Regio, which are shield volcanoes [17, 18, 38].

Table 2 Volcano characteristics

Volcano	Basal Diameter (km)	Height (km)	RMS Slope (degrees)	Average Slope (degrees)	Volume ($x10^4$ km^3)
Sapas Mons	630	2.34	0.646	0.537	10.9
Maat Mons	405	6.8	2.19	1.66	30.5
Ozza Mons	495	3.85	0.719	0.764	42.3
Hawaii	200	8.48	5.60	5.01	12.5
Pico de Fogo	63	6.28	13.55	12.06	1.46

A possible lithospheric model below Sapas Mons is shown in Figure 4. The topographically high Sapas Mons volcano between 1000-2300 km, corresponds to a bouguer low, reflecting a possible lower-density partial melting/magma chamber. This compares to the topographically low plains region surrounding Sapas Mons (0-1000 km) which has a thinner crust (~20 km). A thinner (~70 km) thermal lithosphere (the crust and upper mantle layers) and higher surface heat flow (~38 mW m^{-2}), occurs below Sapas Mons and Atla Regio, which could be associated with the presence of a mantle plume.

Flexural modelling at Atla Regio and Oahu can be observed in Figures 3 & 5. The distance between the half-width of the depression (x) and the maximum amplitude of the forebulge (x_b) were found to be 324 km, 310 km and 212 km for Atla Regio, Oahu and Cape Verde, respectively.

Using Equations [4] and [5] a flexural parameter of 103 km, 98.7 km and 67.5 km and a flexural rigidity of $8.19x10^{23}$ N m, $5.34x10^{23}$ N m and $1.17x10^{23}$ N m were found for Atla Regio, Oahu, and Cape Verde, respectively. Substituting these flexural rigidity estimates into Equation [6], an elastic lithosphere thickness of 50.9 km, 44.1 km and 26.6 km were found for Atla Regio, Oahu and Cape Verde, respectively. The thermal gradient for Atla Regio was calculated from Equation [7] and found to be 6.19 K km^{-1}. Using this thermal gradient with Equation [8] a total heat flux of 20.4 mW m^{-2} for basal lithospheric thermal conductivities of 3.3 W m^{-1} K^{-1} was found.

Since Equation [7] is specific of Venus, the thermal gradient at Oahu was found from a simple oceanic geotherm [12] and yielded a geothermal gradient of 23 K km^{-1} and 20 K km^{-1} for

Oahu and Cape Verde, respectively. This gradient and a thermal conductivity of 2.55 W m^{-1} K^{-1} was used to determine a total heat flux of 59 mW m^{-2} and 51 mW m^{-2} for Oahu and Cape Verde, respectively.

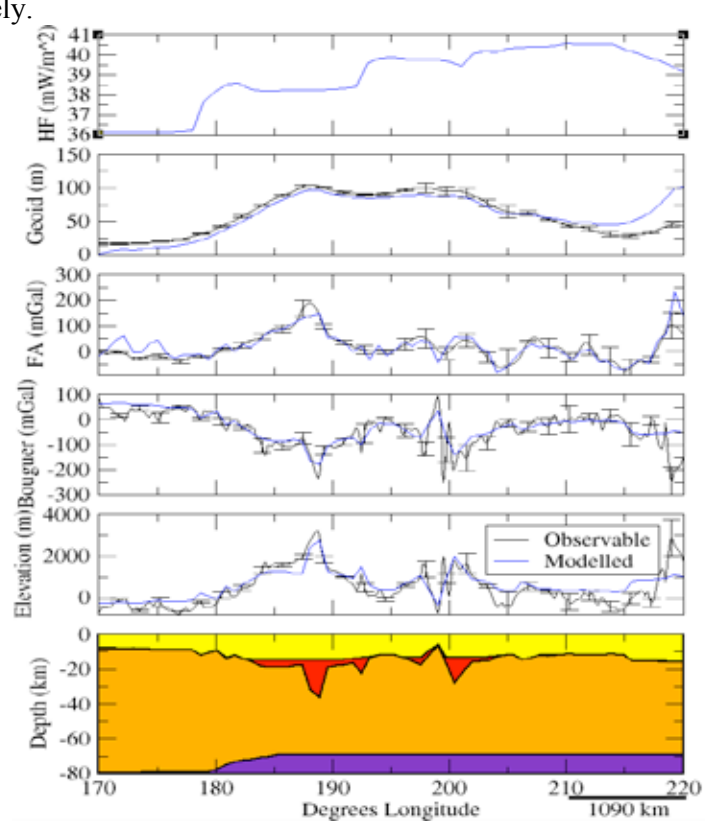


Figure 4 Profile across Sapas Mons at 8° latitude, showing the crust (yellow and red) and conductive upper mantle (orange) that comprise the thermal lithosphere. The adiabatic sub-lithosphere is shown in purple. Note: The plot of heat flow has no observable. Scale bar: 10° Longitude = 1090km

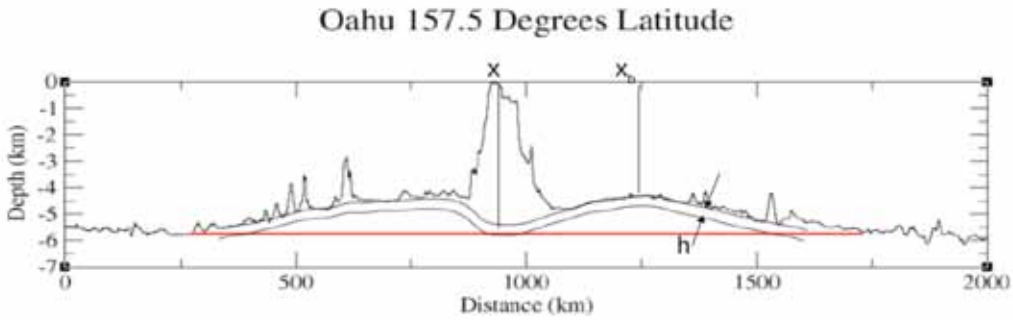


Figure 5 North-South topographic profile across Oahu, Hawaii. The red line is the constant-depth line, facing arrows infer an elastic lithosphere of thickness (h).

Discussion/Conclusion

The excess temperature due to a mantle plume on Venus was found to range from 85-769°C for unconstrained stagnation distances between 200-600 km, compared to the excess temperature of 193°-249°C and 206 K for more constrained stagnation distances of 350-450 km and 390 km for the Hawaiian and Cape Verde swells, respectively. One reason for a large range of excess temperatures at Atla Regio compared to Oahu and Cape Verde is that this area has a larger cross-sectional area ~3056 km^2 compared to Oahu (1188 km^2) and Cape Verde

(1097 km^2).

Due to Venus' lack of water, a higher density contrast between mantle and air also produced higher excess temperatures compared to the smaller density contrast arising from mantle and water on Earth. The equation for excess temperature also relies on a stagnation distance, which is calculated from the radial velocity of the plume, the average unperturbed asthenospheric velocity and the half plate velocity at the hot-spot (Equation 8 in [9]). The stagnation distances for the Hawaiian and Cape Verde swells are also better constrained, than those for Atla Regio, since Venus has a stationary lithosphere, with zero plate velocities. The closest approximation of stagnation distances on Venus may be the diameter of coronae, due to their possible mantle plume origin (200-600 km) [13]. The existence of an asthenospheric channel that extends throughout the entire mantle is questionable on Venus [10], however our calculations of excess temperature assumed an asthenospheric channel thickness. This thickness was more tightly constrained for Hawaii (~100±16 km), compared to Atla Regio (~200 ±100 km) [8, 12].

The buoyancy flux and heat flow values obtained also rely on the velocity of the hot-spot relative to the lithosphere [8, 9]. For Oahu the hot-spot velocity ranged from 83 mm yr^{-1} to 96 mm yr^{-1} [4, 9, 31, 32], with ~90 mm yr^{-1} being chosen in this study, similar to the value used by [12]. This yielded a mantle plume heat flux of 3.25×10^{11} W at Oahu, slightly higher than the mantle plume heat flux of 3×10^{11} W obtained by [12] at a similar location.

Buoyancy flux and heat flow estimates for a mantle plume at Atla Regio, however, have a greater uncertainty than values obtained for Hawaii, since Venus has a stationary lithosphere [1]. The closest Earth-based hot-spot analogue to Venus are those located on slow-moving plates [14]. By taking the lowest estimate of the plate velocity of the Cape Verde hot-spot (12 mm yr^{-1} [27]), and taking the heat flux obtained as an upper bound, heat flux from a mantle plume through Atla Regio should range from 0 W to 1.53×10^{11} W. This compares to the heat flux of 4.0×10^{10} W found for Cape Verde. Even though the same plate velocities were used for Atla Regio and Cape Verde, the differences in cross-sectional area and surface density contrasts caused the variation in heat flow estimates for each of these areas. Despite the high uncertainties and large range of estimated values for excess temperature, buoyancy flux and heat flow, this study has showed that dynamic support from a mantle plume may be contributing to the topography at Atla Regio.

The volcanoes at Atla Regio have larger basal diameters and volumes (except Sapas Mons), but lower elevations (except Maat Mons) than Hawaii and Pico de Fogo. This is consistent with the general trend that volcanoes on Venus have larger basal diameters and volumes than shield volcanoes on Earth [1, 18]. This difference can be attributed to Venus' lack of plates and the associated velocities over hot-spots [1]. On Venus, volcanoes would remain over the hot-spot for longer time intervals compared to Earths' volcanoes, where a plate would move away from the hot-spot over time [1]. Volcanoes within Atla Regio also tend to have shallower slopes than Hawaii and Pico de Fogo, which could be due to: 1) higher temperatures within the crust and surface; 2) lower-viscosity lavas; 3) larger volumes of magma erupted at higher effusion rates; 4) propagation of lavas by lava tubes and; 5) higher surface pressures [18, 19, 20].

The thin thermal lithosphere, high heat flow and presence of a low-density root below Sapas Mons could reflect current volcanism associated with an upwelling mantle plume. The excess heat associated with a mantle plume would act to thin the thermal lithosphere, producing a geoid high, and causing partial melting below this area. The associated bouguer low could represent either this partial melt or a possible low-density magma chamber.

The elastic lithosphere thickness of 50.9 km, obtained at Atla Regio from flexural modelling, is close to the upper limit of the range (40-50 km) of elastic lithosphere thicknesses from Phillips [15], who used spectral analysis and Monte Carlo inversions. The elastic lithosphere thickness obtained at Oahu (44.1 km) is higher than that obtained by Turcotte & Schubert [12] over the Hawaiian Archipelago (34 km). The elastic lithosphere thickness at the Cape Verde Archipelago was found to be 26.6 km, within the range found by [6] from spectral admittance studies, but lower than the 29 km found from best-fit surface/subsurface loading models. These differences in estimates of elastic lithosphere thicknesses may be caused from the difficulties associated with determining the maximum amplitude of the forebulge due to interference from the surrounding topography that acts to obscure the flexural upwarp, similar to what is observed on the left-side of the profile in Figure 4.

The thermal gradient of 6.19 K km^{-1} for Atla Regio, obtained indirectly from flexural modelling, was also within the range of thermal gradients [7-10 K km^{-1}] outlined by Phillips [15]. Phillips [15] suggested Atla Regio should have a total heat flux less than what is obtained from Earth-scaling (~80 mW m^{-2}). From flexural modelling a total heat flux of 18.6-24.8 mW m^{-2} was found for Atla Regio, which is significantly lower than that obtained by Earth scaling. The total heat flux value obtained from flexural modelling at Oahu is slightly higher (58.7 mW m^{-2}) than those obtained from total heat flux measurements taken at Hawaii (52.9±2.5 mW m^{-2} [26]) , whilst the obtained heat flux value for Cape Verde is is slightly less (51 mW m^{-2}) than measured values (61±7.4 mW m^{-2} [27]). Differences in the heat flow estimates obtained from flexural modelling to measured values are most likely caused from comparing these areas to a globally averaged oceanic geotherm.

The results of this study suggests that volcanic topography at Atla Regio could be produced by three processes, thermal, volcanic and flexural. The impingement of a mantle plume on the lithosphere could have initiated rifting and volcanism, resulting in volcanic loading and flexure of the lithosphere. These processes act in a similar way to those observed on Earth, but yield different results due to Venus' unique surface conditions. Future work aims to constrain the stagnation distances through finite-element modelling and to produce a plausible rheological structure below Atla Regio, which will be used to constrain these processes further.

Acknowledgments

The authors would like to thank the Space Policy Unit, Engineers Australia, and ASRI for providing sponsorship to attend the 10th Australian Space Science Conference in Brisbane. The authors would like to thank Dr Marc Norman and Dr Jon Clarke for their time and suggestions for improvements to this paper. Software used in this study is available at the following sites , SHTOOLS: http://www.ipgp.fr/~wieczor/SHTOOLS/SHTOOLS.html , GMT: http://gmt.soest.hawaii.edu/, Grace: ttp://plasma-gate.weizmann.ac.il/Grace/, LitMod3D: http://www.es.mq.edu.au/staff/Software1.htm). This is contribution 694 from the Australian Research Council National Key Centre for the Geochemical Evolution and Metallogeny of Continents (http://www.gemoc.mq.edu.au).

References

1. Moore, P. "*Venus*". Cassel Illustrated, Great Britain, 2002, pp. 191.
2. Senske, D.A. & Head, J.W. "Atla Regio Venus: Geology and origin of a major equatorial volcanic rise. *Proceedings of Lunar and Planetary Science*, No. 789, 1992, pp.107-109.
3. Phillips, R.J., Grimm, R.E., & Malin, M.C. "Hot-spot evolution and the global

tectonics of Venus". *Science*, Vol. 252, No. 5006, 1991, pp. 551-658.
4. Davies, G.F. "Ocean bathymetry and mantle convection: 1. Large-scale flow and hotspots". *Journal of Geophysical Research*, Vol. 93, No. B9, 1988, pp. 10467-10480.
5. Wilson, J.T. "A possible origin of the Hawaiian Islands". *Canadian Journal of Physics*. Vol. 41, 1963, pp. 863-870.
6. Ali, M.Y., Watts, A.B., & Hill, I. "A seismic reflection profile study of lithospheric flexure in the vicinity of the Cape Verde Islands". *Journal of Geophysical Research*, Vol. 108, No. B5, 2003, pp. 24.
7. Morgan, W.J. "Convection plumes in the lower mantle". *Nature*, Vol. 230, No. 42, 1971, pp. 42-43.
8. Sleep, N.H. "Hotspots and mantle plumes: some phenomenology". *Journal of Geophysical Research*, Vol. 95, No. B5, 1990, pp. 6715-6736.
9. Sleep, N.H. "Hotspot volcanism and mantle plumes". *Annual Review of Earth and Planetary Sciences*, Vol. 20, 1992, pp. 19-43.
10. Jellinek, M.A., Lenardic, A. and Manga, M. "The influence of interior mantle temperature on the structure of plumes: Heads for Venus, Tails for Earth". *Geophysical Research Letters*, Vol. 29, 2002, pp. 4.
11. Schubert, B., Bercovici, D, Glatzmaier, G.A. "Mantle dynamics in Mars and Venus: Influence in an immobile lithosphere on three-dimensional mantle convection". *Journal of Geophysical Research,* Vol. 95, No. B9, 1990, pp. 14105-14129.
12. Turcotte, D.L. & Schubert, G. "*Geodynamics*" 2nd Ed, Cambridge University Press, 2002, pp. 458.
13. Head, J.W., Crumpler, L.S., Aubele, J.C., Guest, J.E., Saunders, R.S. "Venus volcanism: Classification of volcanic features and structures, associations and global distribution from Magellan data". *Journal of Geophysical Research,* Vol. 97, No. E8, 1992, pp. 13153-13197.
14. Phillips, R.J., Johnson, C.L., Mackwell, S.J., Morgan, P., Sandwell, D.T., & Zuber, M.T. "Lithospheric mechanics and dynamics of Venus". In *Venus II: Geology, Geophysics, Atmosphere and solar wind environment*, Bougher, S.W., Hunten, D.M. & Phillips, R.J. (eds.), 1997, The University of Arizona Press, pp. 1362.
15. Phillips, R.J. "Estimating lithospheric properties at Atla Regio, Venus". *Icarus*, Vol. 112, 1994, pp. 147-170.
16. Smrekar, S., & Parmentier, E.M. "The interaction of mantle plumes with surface thermal and chemical boundary layers: applications to hotspots on Venus". *Journal of Geophysical Research*, Vol. 101, 1996, pp. 5397-5410.
17. Skinner, B., Porter, S.C. & Park, J. "*Dynamic Earth: An introduction to physical geology"* 5th Ed, Wiley, 2004, United States of America.
18. Schaber, G.G. "Volcanism on Venus as inferred from the morphometry of large shields". *Proceedings of Lunar and Planetary Science*, Vol. 21, 1991, pp. 3-11.
19. Head, J.W. & Wilson, L. "Volcanic processes and landforms on Venus: Theory, prediction and observations". *Journal of Geophysical Research*, Vol. 91, 1986, pp. 9407-9446.
20. Wood, C.A. "Venusian volcanism: Environmental effects on style and landforms". In *Reports of the Planetary Geology and Geophysics Programs-1979*, 1979, pp. 244-246.
21. Fullea, J., Afonso, J.C., Connolly, J.A.D., Fernandez, M., Garcia-Castellanos, D., & Zeyen, H. "LitMod3D: An interactive 3D software to model the thermal, compositional, density, seismological, and rheological structure of the lithosphere and sublithospheric upper mantle". *Geochemistry, Geophysics, Geosystems*, Vol. 10, No. 8, 2009, pp. 1-21.
22. Afonso, J.C., Fernandez, M., Ranalli, G., Griffin, W.L., & Connonlly, J.A.D. "Integrated geophysical-petrological modeling of the lithosphere and sublithospheric

upper mantle: Methodology and applications". *Geochemistry, Geophysics, Geosystems*, Vol. 9, No. 5, 2008, pp. 1-36.
23. Barnett, D.N., Nimmo, F., & McKenzie, D. "Flexure of Venusian lithosphere measured from residual topography and gravity". *Journal of Geophysical Research,* Vol. 107, 2002, pp. 1-21.
24. Smrekar, S.E. "Evidence for active hotspots on Venus from analysis of Magellan gravity data". *Icarus*, Vol. 112, 1994, pp. 2-26.
25. Ebinger, C.J., Bechtel, T.D., Forsyth, D.W., and Bowin, C.O. "Effective elastic plate thickness beneath the East African and Afar Plateaus and dynamic compensation of the uplifts". *Journal of Geophysical Research*, Vol. 94, 1989, pp. 2883-2901.
26. Von Herzen, R.P., Detrick, R.S., Crough, S.T., Epp, D. & Fehn, U. "Thermal origin of the Hawaiian Swell: Heat flow evidence and thermal models". *Journal of Geophysical Research*, Vol. 87, 1982, pp. 6711-6723.
27. Courtney, R.C. & White, R.S. "Anomalous heat flow and geoid across the Cape Verde Rise: Evidence for dynamic support from a thermal plume in the mantle". *Geophysical Journal Royal Astronomical Society*, Vol. 87, 1986, pp. 815-868.
28. Head, J.W. & Wilson, L. "Magma reservoirs and neutral buoyancy zones on Venus: Implications for the formation and evolution of volcanic landforms". *Journal of Geophysical Research,* Vol. 97, 1992, pp. 3877-3903.
29. Seiff, A. "Thermal structure of the atmosphere of Venus". In *Venus,* Hunten, D.M., Colin, L., Donahue, T.M., & Moroz, V.I. (eds.), 1983, The University of Arizona Press, pp. 215-279.
30. Johnson, C. & Sandwell, D.T. "Lithospheric flexure on Venus". *Geophysical Journal International*, Vol. 119, 1994, pp. 627-647.
31. Gordon, R.G. & Jurdy, D.M. "Cenozoic global plate motions". *Journal of Geophysical Research*, Vol. 91, 1986, pp. 12389-12406.
32. Duncan, R.A. & Clague, D.A. "Pacific plate motion recorded by linear volcanic chains". In *The ocean and basins and margins*. *Vol. 7A. The Pacific Ocean* (Naim, A.E.M., Stehi, F.G., Uyeda, S. eds.) 1985, New York, pp. 89-121.
33. Fullea, J., *"Development of numerical methods to determine the lithospheric structure combining geopotenital, lithostatic, and heat transport equations. Applications to the Gibraltar Arc System"*. PhD thesis, University of Barcelona, Spain, 2008, pp. 240.
34. Horai, K. "Thermal conductivity of Hawaiian basalt: A new interpretation of Robertson and Peck's data". *Journal of Geophysical Research*, Vol. 96, No. B3, 1991, pp. 4125-4132.
35. Robertson, E.C. & Peck, D.L. "Thermal conductivities of vesicular basalt from Hawaii". *Journal of Geophysical Research,* Vol. 79, No. 32, pp. 4875-4888.
36. Shepard, M.K., Campbell, B.A., Bulmer, M.H., Farr, T.G., Gaddis, L.R., & Plaut, J.J. "The roughness of natural terrain: A planetary and remote sensing perspective". *Journal of Geophysical Research*. Vol. 106, No. E12, 2001, pp. 32777-32795.
37. Kreslavsky, M.A. & Head III, J.W. "Kilometer-scale slopes on Mars and their correlation with geologic units: Initial results from Mars Orbiter Laser Altimeter (MOLA) data". *Journal of Geophysical Research*, Vol. 104, 1999, pp. 21911-21924.
38. Day, S.J., Heleno, de Silva, S.I.N., Fonseca, J.F.B.D. "A past giant lateral collapse and present-day flank instability of Fogo, Cape Verde Islands". *Journal of Volcanology and Geothermal Research*, Vol. 94, 1999, pp. 191-218.

An Initial Model Assessment of NO_x Photochemistry on Venus with Heterogeneous Oxidation of CO

M. Shunmuga Sundaram*†, F. P. Mills‡§, M. Allen¶‖ and Y. L. Yung‖

* *Research School of Physical Sciences and Engineering, Australian National University, Canberra, ACT, Australia, 2601*
† *Now at Centre for Civil Society, India,*
‡ *Research School of Physics and Engineering, Australian National University, Canberra, ACT, Australia, 2601*
§ *Fenner School of Environment and Society, Australian National University, Canberra, ACT, Australia, 2601*
¶ *Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA*
‖ *Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA 91125, USA*

Summary: The observational discovery of NO in the Venus cloud layers has revived interest in odd nitrogen chemistry on Venus. The discovery paper suggested, based on column-integrated results, that odd nitrogen chemistry might be important for understanding the dominant chemical cycles on Venus. To make a more detailed assessment of the potential impacts of odd nitrogen chemistry, a numerical model with heterogeneous oxidation of CO, which gives a column abundance for O_2 near the observational upper limit, was used to simulate the chemistry in Venus' mesosphere. The model simulations suggest odd nitrogen chemistry may have significant effects on specific aspects of the modelled chemistry, such as SO and SO_2 abundances at 80–90 km altitude. However, the global-scale impact of odd nitrogen chemistry appears to be small.

Keywords: Venus, nitrogen chemistry, mesosphere, heterogeneous chemistry

Introduction

One of the most puzzling unresolved questions for planetary atmospheres is what process(es) is (are) responsible for the chemical stability of CO_2 on Venus. Heterogeneous oxidation of CO can reduce the modelled O_2 abundance to near its observational upper limit [1] but probably distorts the CO profile outside the observational bounds [2]. Nitrogen oxide chemistry is known to modify catalytic chemistry cycles, [3] and the impact of nitrogen oxide chemistry on the catalytic chemistry in Venus' atmosphere has not been examined in detail [4]. This modelling study examines the effects of introducing nitrogen oxide chemistry into a Venus atmospheric chemistry model in which heterogeneous oxidation of CO is an important pathway for production of CO_2.

Venus' atmosphere is composed primarily of CO_2 (96.5%) and N_2 (3.5%) [5], with trace abundances of sulfur, chlorine, and hydrogen oxide compounds [5]–[7]. The three dominant chemical cycles are believed to be the CO_2 cycle, the sulfur oxidation cycle, and the polysulfur cycle. The CO_2 cycle involves photodissociation of CO_2 on the day side at wavelengths, λ,

$\lesssim 200$ nm to produce CO and O. A substantial portion of the CO and atomic oxygen produced on the day side is transported to the night side where the atomic oxygen combines via Reaction 1 to form O_2 in a variety of excited states.

$$2O + M \rightarrow O_2^* + M \tag{1}$$

where M is any third atom or molecule that can collide with and stabilise the intermediate complex and O_2^* is an O_2 molecule in an excited (electronic) state. Evidence for a fast rate of production of O_2 on the night side comes from observations of intense airglow emission, Reaction 2, on the night side [8]–[10].

$$O_2(a^1\Delta) \rightarrow O_2(X^3\Sigma) + h\nu(\sim 1.27\mu m) \tag{2}$$

CO_2 is then produced from CO and O_2 at a rate that balances its loss. Simulations have found that if a planetary atmosphere were initially pure CO_2, then CO and O_2 would accumulate to be about 7 and 3.5 %, respectively [11], [12]. The simulated accumulations of CO and O_2 are large in this case because the reaction rate coefficient for direct recombination of CO and O to form CO_2 is small, several orders of magnitude smaller than that for Reaction 1 at the temperature typically found in planetary atmospheres. The observed abundance of CO in Venus' atmosphere, however, is much smaller, $\sim 20 - 50$ ppm and the abundance of O_2 has not been quantified because it has been below the detection limit on every attempt to measure it [13]–[15]. The most restrictive interpretation [16] of the most sensitive observation [14] is equivalent to a column-averaged, uniform (volume) mixing ratio of 1.6×10^{-7}. Although there are disagreements over interpretation of the upper limit observations [2], the upper limit on O_2 is much smaller than the observed abundance of CO.

In a more realistic planetary atmospheric model, production of CO_2 occurs primarily via catalytic processes, such as Reaction 3.

$$\begin{array}{rcl} Cl + CO + M & \rightarrow & ClCO + M \\ ClCO + O_2 + M & \rightarrow & ClCO(O)O + M \\ ClCO(O)O + O & \rightarrow & Cl + CO_2 + O_2 \\ net: \ CO + O & \rightarrow & CO_2 \end{array} \tag{3}$$

Even with the incorporation of such powerful catalytic processes, modelling studies using only gas-phase chemistry have had great difficulty oxidising CO at a rate that is sufficiently fast to reduce the modelled O_2 abundance below the observational upper limit [2], [7]. Simulations including heterogeneous oxidation of CO, Reaction 4, on the particles comprising Venus' global sulfuric acid clouds, however, have shown the modelled O_2 abundance can be reduced to a level comparable to the observational upper limit if the reactive uptake coefficient, γ_{rxn}, is $\gtrsim 10^{-5}$ [1], [17].

$$CO + O + aerosol \rightarrow CO_2 + aerosol \tag{4}$$

Heterogeneous reactions are reactions that involve multiple phases, such as liquid and gas. Such reactions have been found to be extremely important in the terrestrial atmosphere and are a key factor in the production of the Antarctic "ozone hole" [18]. The rate at which these reactions occur is parameterised in terms of a reactive uptake coefficient, γ_{rxn}, which incorporates diffusion, solubility, and reactivity effects. Limited experimental evidence suggests CO may be oxidised on sulfuric acid to CO_2, [19]–[21] but the studies to date have elucidated neither the mechanism by which CO is oxidised nor the rate at which the oxidation occurs.

N_2 is the second most abundant species in Venus' atmosphere, but it is largely unreactive unless the strong N-N bond is broken to form odd nitrogen (eg., N, NO, NO_2), also known as NO_x. This occurs via extreme ultraviolet photodissociation at high altitudes and leads to NO airglow emission, which is observed on the night side of Venus [22]. At lower altitudes, lightning and cosmic rays can split the N-N bond to produce odd nitrogen. Cosmic rays are highly energetic (~ 100 MeV-5 GeV) particles, primarily protons and alpha particles, that originate from the Sun and other stars in the galaxy. Depending on energy, they can penetrate deeply into a planetary atmosphere and ionise or split molecules, including N_2. The rate of production of NO_x by cosmic rays in Venus' atmosphere is estimated to be small [4], but potentially significant depending on the intensity and frequency of lightning. Lightning releases a tremendous amount of energy in a localised region and is able to drive a range of chemical reactions, including localised ionisation of molecules. The initial effects from lightning are damped as the release of energy dissipates. Consequently, the net chemical effects from lightning are equivalent to those from thermodynamic equilibrium chemistry at about 2000 K [7], and may be summarised by Reaction 5 [23].

$$\begin{aligned} CO_2 &\leftrightarrow CO + O \\ O + N_2 &\leftrightarrow NO + N \\ N + CO_2 &\leftrightarrow NO + CO \\ net: \; 2CO_2 + N_2 &\leftrightarrow 2CO + 2NO \end{aligned} \tag{5}$$

Nitrogen oxide chemistry was included in an early photochemical model for Venus [3], but it has been neglected in more recent Venusian photochemical models, which have concentrated on the chlorine [24] and sulfur [2] aspects of Venusian atmospheric chemistry. One reason is that the importance of NO_x chemistry on Venus has been a subject for debate due to conflicting evidence over the intensity and frequency of lightning. Attempts to detect the optical signature of lightning have consistently indicated very low rates for intense lightning [25]. Measurements of electromagnetic waves by magnetometers on spacecraft orbiting Venus, however, have repeatedly identified waves whose origin is believed to be uniquely tied to lightning [26]. The reported detection of NO below 60 km altitude via observation of two infrared absorption lines [4] appears to have established the existence of reactive NO_x in the cloud layers on Venus. The abundance reported, 5.5 ± 1.5 ppb [4], appears to be too large for it to be produced by cosmic rays, which implies the release of energy via lightning on Venus is sufficiently large to produce the observed amount of NO [4], [27]. The reported detection of NO was based on a single night's observation of a relatively small region of Venus. Consequently, a significant observational sampling uncertainty exists in the reported detection and the time- and spatial-average abundance of NO could be a factor of several smaller or larger than that reported.

Direct oxidation of CO to CO_2 via, for example, reaction with NO_2 has neither been

observed in laboratory studies nor proposed in previous modelling studies. However, based on analogies to terrestrial atmospheric chemistry, odd nitrogen may play a range of (direct and indirect) roles in the oxidation of CO and SO_2 and may help convert unreactive HCl into more reactive Cl via $ClNO_2$ and $ClONO_2$. Its importance to Venus' overall chemistry is highly dependent on its abundance and vertical distribution.

An initial assessment of the impact of NO_x chemistry on Venus has been made using a simplified model [4], which was based, in turn, on a nominal chemistry model for Venus [12]. Calculations using the simplified model suggest the reported NO abundance could have an important effect on the modelled photochemistry [4]. The two most significant effects were found to be oxidation of SO via Reaction 6 and production of O_2 via Reaction 7 [4].

$$NO_2 + SO \rightarrow NO + SO_2 \tag{6}$$

$$NO_2 + O \rightarrow NO + O_2 \tag{7}$$

Reaction 6 was assessed to return half of the SO_2 that was photodissociated, and Reaction 7 provided 13% of the simplified model's total O_2 production [4]. These results from the simplified model need to be assessed in the context of a more complete Venusian photochemical model as changes in other areas of the model may either reduce or enhance the first-order effects determined from the simplified model. In addition, the column-integrated O_2 abundance in the nominal Venus chemistry model on which the simplified model was based is an order of magnitude larger than the observational upper limit. This may have important implications for the overall impact of NO_x chemistry, which also must be examined within a more comprehensive photochemical model.

This paper examines the impact of adding NO_x chemistry to a comprehensive Venusian mesospheric photochemical model. Heterogeneous oxidation of CO via Reaction 4 is used to reduce the simulated column abundance of O_2 to be near the observational upper limit. The observational sampling uncertainty associated with the reported detection of NO is assessed by varying the lower boundary condition for NO in the model over a range of values. The Venusian mesospheric photochemical model and the NO_x chemistry that was added to it are described in the next section. Results from calculations over a range of NO values are then presented, and the impacts from adding the NO_x chemistry are assessed.

Photochemical Model

Details on the basic photochemical model have been published [1], [2], [12], [15], [24], [28], so only a summary description [2] is provided here. The basic Venus photochemical model used for this study is customised from the Caltech/JPL code [29]. This code solves the continuity equation simultaneously for each species in the model and was also used to study the photochemistry of Mars [11]. The Venus model is one-dimensional and steady-state, with vertical transport between the layers. It is converged to a steady-state solution via a finite-difference iterative algorithm. The model atmosphere, noon local solar time [30], extends from 58 to 112 km altitude with 2-km thick layers. The model represents the global average

photochemistry, so diurnal variations have not been explicitly modeled. Photodissociation was calculated for 45 degrees latitude at local noon and then divided by two. The solar fluxes used represent high solar irradiance conditions. Diffuse radiation is calculated using the Feautrier method generalised to arbitrary anisotropic scattering with aerosol scattering properties based on Pioneer Venus measurements [31].

Photoabsorption and chemical kinetic data were drawn from critically reviewed compilations (e.g., [32]) whenever possible. Changes in kinetic rate recommendations since 1997 are not expected to be significant for the basic Venus model. The nominal equilibrium constant for ClCO formation and decomposition, $K_{ClCO} = 1.6 \times 10^{-25} exp(4000/T)$ [33], is based on laboratory work by Nicovich et al [34]. Temperature dependent cross sections were used for CO_2 [35], [36], SO_2 [37]–[44], O_3 [45]–[52], and OCS [53]. UV cross sections for photolysis of ClC(O)OO were included [24].

Vertical transport via eddy diffusion was set based on observations [54]–[56]. The H_2O profile was fixed to match the equilibrium vapor pressure over 75 wt% sulfuric acid [28], which is still commonly assumed as the composition of the upper cloud layer and overlying haze although Venera measurements indicate the composition is 85 wt% [16]. At the lower boundary, the mixing ratio for CO_2 was set to 0.965 [5], HCl to 0.4 ppm [57], [58], OCS to 1 ppb (as was assumed by Bezard et al [59]), and SO_2 to 1 ppm. The concentration gradient at the lower boundary for all other species was set to zero. At the upper boundary, the upward flux of molecules, such as CO_2, was set equal to the downward flux of their photodissociation products to simulate photodissociation occurring above the upper boundary.

Thermal decomposition of ClCO was decreased by a factor of 1.5 times the assessed uncertainty in the equilibrium constant for ClCO formation and decomposition. This significantly increases the efficiency of CO oxidation via chlorine catalytic chemistry, particularly at about $80 - 90$ km altitude.

Heterogeneous oxidation of CO on aerosols was included with a reactive uptake coefficient, $\gamma_{rxn} = 10^{-5}$. This means Reaction 4 is a primary pathway for oxidation of CO when integrated over all modelled altitudes [1], [17].

The nitrogen oxide chemistry introduced in this initial model assessment [3], Table 1, includes 26 reactions and 7 species. Most rate coefficients and cross sections were taken from the latest stratospheric chemistry assessment [60]. Coupling between nitrogen oxide and chlorine oxide chemistry via chlorine nitrate was not included in this study. Oxidation of SO_2 to SO_3 by NO_x also was not included. The photodissociation reactions in Table 1 were all calculated using room temperature cross sections for this initial assessment. Yung and DeMore used a similar set of reactions [3], while Krasnopolsky used a simplified set of four reactions [4], three of which are Reactions 116, 444, and 446+447 in this study. The fourth reaction he included was Reaction 8 [4], which is important at higher altitudes but not important for the lower altitude chemistry of interest in this study.

$$N + O \rightarrow NO + h\nu \tag{8}$$

Large changes have occurred in about one-third of the reaction rate coefficients since the modelling by Yung and DeMore [3]. The rate coefficient used for Reaction 440 in this study,

which was taken from an earlier study [3], is a factor of 1.5-2 slower than that used by Krasnopolsky [4]. This difference appears to be due to the greater efficacy of CO_2 as a third-body for stabilising the intermediate complex [61], [62]. The rate coefficient used for Reaction 444 in this study, determined via laboratory measurements [63], is 30% slower at 300 K than that used by Krasnoposky [4] and a factor of 2 faster than his at 200 K. The rate coefficient used by Krasnopolsky for Reaction 444 is based on fitting photochemical model calculations to measured profiles of NO in the Martian and terrestrial thermospheres [64].

Production of odd nitrogen via lightning was not included in the reaction set. Instead, nitrogen oxide chemistry was simulated by specifying the mixing ratio for NO at the 58 km altitude lower boundary of the model. The lower boundary mixing ratio was varied from the nominal observed value of 5.5 ppb to 30 ppb to assess the potential impact of the observational sampling uncertainty.

Model Results

A detailed analysis of the full implications of the calculations is beyond the scope of the present manuscript, so this section and the next will focus on selected highlights.

NO is the most abundant form of odd nitrogen and its mixing ratio decreases slowly with increasing altitude until about 77 km, Fig. 1. At higher altitudes, the optical depth of CO_2, which absorbs over the same wavelength range as NO, decreases and NO photodissociation increases. This increases production of N, which leads to an increasing rate of loss of odd nitrogen via Reaction 444. The largest mixing ratios for NO_2 occur near 85 km altitude where the photodissociation rate for CO_2 is a maximum.

The largest changes in SO and SO_2, Fig. 2, occur where the NO_2 mixing ratio is largest, near 85 km. (The calculated SO and SO_2 profiles with no odd nitrogen are shown in Fig. 3 along with the abundances inferred from selected observations.) The direct effect of the modelled odd nitrogen chemistry is to oxidise SO to SO_2 via Reaction 459. This also increases the modelled abundances of SO because photodissociation of SO_2 to form SO is very rapid at these altitudes. Reaction 459 acts, in effect, to reduce the rate of loss of SO_x (= SO + SO_2) via production of S_8 and SO_3 at these altitudes. Below 65 km, which is within the upper cloud layer, the modelled odd nitrogen chemistry causes SO and SO_2 abundances to decrease. At these altitudes, the modelled loss of SO_2 is likely occurring via an enhancement in the rate of production of $ClSO_2$.

The introduction of odd nitrogen chemistry also has altitude-dependent effects on the abundances of O, O_2, and CO, Fig.4. (The calculated O and O_2 profiles with no odd nitrogen are shown in Fig. 3 along with the abundances inferred from selected observations.) The odd nitrogen chemistry has the net effect of moderating the extrema in O_2 because O_2 has maximum abundance at 75–85 km and minimum abundance within the upper cloud layer in models with strong heterogeneous oxidation of CO. Although the fractional changes in O, O_2, and CO are large in the upper cloud layer, the absolute magnitude of the changes is much smaller, as illustrated in Fig. 5 for CO.

Discussion and Comparisons to Observations

The results from this initial assessment are broadly consistent with those published previously [4]. Odd nitrogen chemistry is important for specific aspects of the Venus mesosphere

Table 1: Nitrogen oxide chemical reactions included in calculations.

Number	Reaction	Rate Coefficient†	Ref
116	$NO + h\nu \rightarrow N + O$	$J_{116} = 2.6 \times 10^{-12}$	a
117	$NO_2 + h\nu \rightarrow NO + O$	$J_{117} = 1.8 \times 10^{-2}$	b§
118	$NO_3 + h\nu \rightarrow NO_2 + O$	$J_{118} = 3.3 \times 10^{-1}$	b§
119	$NO_3 + h\nu \rightarrow NO + O_2$	$J_{119} = 1.1 \times 10^{-1}$	b§
121	$HNO_2 + h\nu \rightarrow OH + NO$	$J_{121} = 3.0 \times 10^{-3}$	b§
122	$HNO_3 + h\nu \rightarrow NO_2 + OH$	$J_{122} = 9.0 \times 10^{-5}$	b§
440	$N + O + M \rightarrow NO + M$	$k_{440} = 1.9 \times 10^{-31} T^{-0.5}$	g
441	$N + O_2 \rightarrow NO + O$	$k_{441} = 1.5 \times 10^{-11} e^{(-3600/T)}$	b
442	$N + O_3 \rightarrow NO + O_2$	$k_{442} = 2.0 \times 10^{-16}$	b‡
443	$N + OH \rightarrow NO + H$	$k_{443} = 5.3 \times 10^{-11}$	c
444	$N + NO \rightarrow N_2 + O$	$k_{444} = 2.1 \times 10^{-11} e^{(100/T)}$	h
446	$NO + O + N_2 \rightarrow NO_2 + N_2$	$k_{446}\ k_0 = 9.0 \times 10^{-32} (T/300)^{-1.5}$	b
		$k_\infty = 3.0 \times 10^{-11}$	b
447	$NO + O + CO_2 \rightarrow NO_2 + CO_2$	$k_{447}\ k_0 = 2.97 \times 10^{-31} (T/300)^{-1.5}$	b∗
448	$NO + O_3 \rightarrow NO_2 + O_2$	$k_{448} = 3.0 \times 10^{-12} e^{(-1500/T)}$	b
449	$NO + H + M \rightarrow HNO + M$	$k_{449} = 1.5 \times 10^{-32} e^{(300/T)}$	d
450	$NO + OH + N_2 \rightarrow HNO_2 + N_2$	$k_{450} = 7.0 \times 10^{-31} (T/300)^{-2.6}$	b
		$k_\infty = 3.6 \times 10^{-11} (T/300)^{-0.1}$	b
451	$NO + OH + CO_2 \rightarrow HNO_2 + CO_2$	$k_{451} = 2.31 \times 10^{-30} (T/300)^{-2.6}$	b∗
452	$NO + HO_2 \rightarrow NO_2 + OH$	$k_{452} = 3.5 \times 10^{-12} e^{(250/T)}$	c
453	$NO + ClO \rightarrow NO_2 + Cl$	$k_{453} = 6.4 \times 10^{-12} e^{(290/T)}$	b
454	$NO_2 + O \rightarrow NO + O_2$	$k_{454} = 5.1 \times 10^{-12} e^{(210/T)}$	b
455	$NO_2 + O + N_2 \rightarrow NO_3 + N_2$	$k_{455} = 2.5 \times 10^{-31} (T/300)^{-1.8}$	b
		$k_\infty = 2.2 \times 10^{-11} (T/300)^{-0.7}$	b
456	$NO_2 + O + CO_2 \rightarrow NO_3 + CO_2$	$k_{456} = 8.25 \times 10^{-31} (T/300)^{-1.8}$	b∗
457	$NO_2 + OH + N_2 \rightarrow HNO_3 + N_2$	$k_{457} = 1.8 \times 10^{-30} (T/300)^{-3.0}$	b
		$k_\infty = 2.8 \times 10^{-11}$	b
458	$NO_2 + OH + CO_2 \rightarrow HNO_3 + CO_2$	$k_{458} = 5.94 \times 10^{-30} (T/300)^{-3.0}$	b∗
459	$NO_2 + SO \rightarrow NO + SO_2$	$k_{459} = 1.4 \times 10^{-11}$	e
460	$HNO + O \rightarrow HNO + O$	$k_{460} = 1.0 \times 10^{-13}$	f
461	$HNO + H \rightarrow H_2 + NO$	$k_{461} = 1.0 \times 10^{-13}$	d
462	$HNO + Cl \rightarrow HCl + NO$	$k_{462} = 1.0 \times 10^{-13}$	f
463	$2HNO \rightarrow N_2O + H_2O$	$k_{463} = 4.0 \times 10^{-15}$	d
464	$HNO_2 + OH \rightarrow H_2O + NO_2$	$k_{464} = 6.6 \times 10^{-12}$	d

† Units for photodissociation (J) and two-body and three-body reaction (k) rate coefficients are sec^{-1}, $cm^3\ sec^{-1}$, and $cm^6\ sec^{-1}$, respectively. Numerical values for photodissociation refer to 70 km, mid-latitude. § Photodissociation of NO_2, NO_3, HNO_2, and HNO_3 was calculated using room temperature cross sections. ‡ Upper limit. ∗ Rate coefficient for M=N_2 multiplied by 3.3 based on mean ratio for association reactions with $M = N_2$ and CO_2 [12]. [a] [4]. [b] [60]. [c] [3]. [d] [65]. [e] [66]. [f] Estimated [3]. [g] [3], [22]. [h] [3], [22], [63].

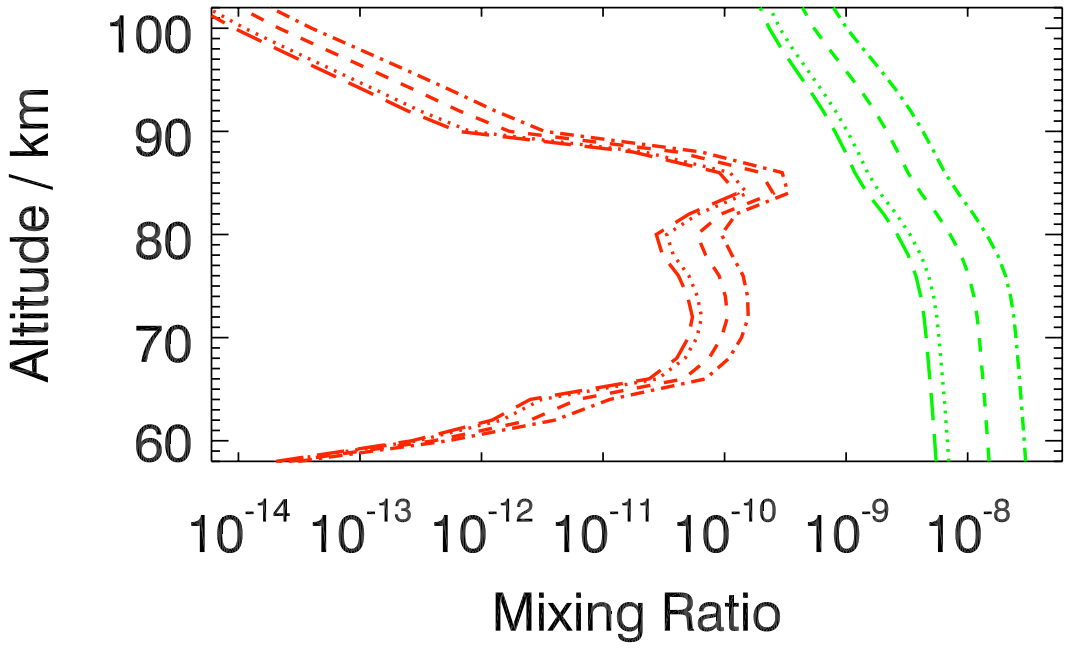


Fig. 1: NO (green) and NO_2 (red) mixing ratios for a range of lower boundary conditions – NO(58 km) = 5.5 ppb (long-dash) as reported from observations [4], NO(58 km) = 7.0 ppb (dotted), NO(58 km) = 15 ppb (short-dash), and NO(58 km) = 30 ppb (dash-dot).

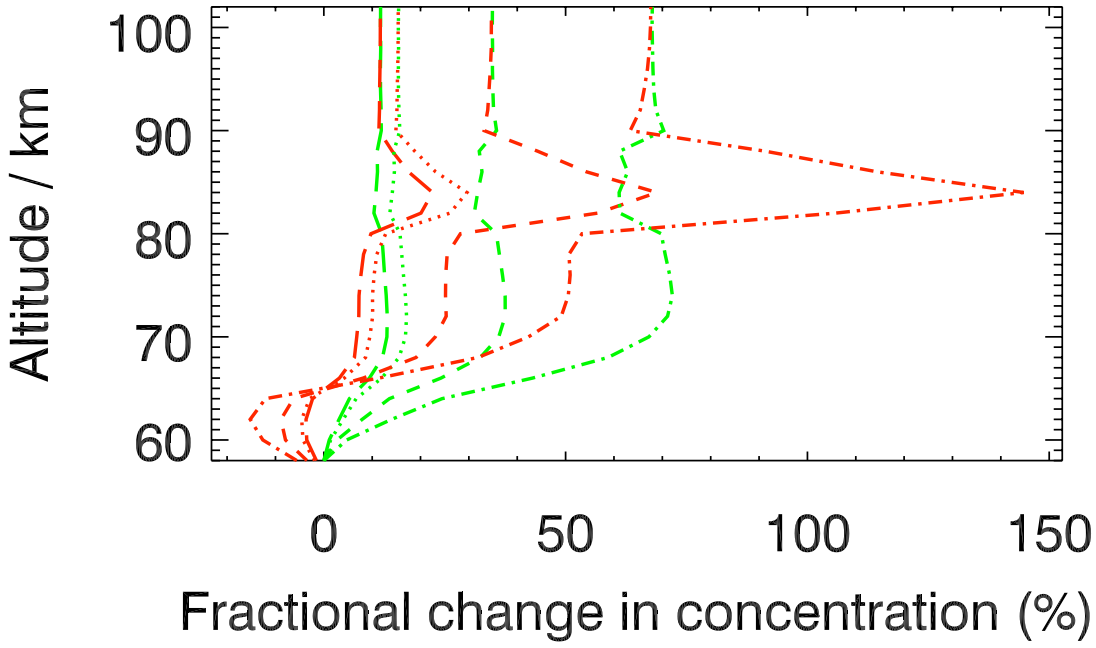


Fig. 2: Fractional difference in SO_2 (green) and SO (red) for a range of lower boundary conditions – NO(58 km) = 5.5 ppb (long-dash) as reported from observations [4], NO(58 km) = 7.0 ppb (dotted), NO(58 km) = 15 ppb (short-dash), and NO(58 km) = 30 ppb (dash-dot). The reference calculation has no NO_x but otherwise has identical input.

model. The effect on SO and SO_2 abundances at 80–90 km, for example, suggests odd nitrogen chemistry may need to be considered when interpreting altitude-resolved observations of SO and SO_2 [67]. However, even in this case, the impact of odd nitrogen chemistry is relatively small for the observed 5.5 ppb abundance of NO below 60 km. Further, odd nitrogen chemistry does not appear to have a substantial impact on global-scale measures, such as the column abundance of O_2, S_8 production, or H_2SO_4 production, Table 2.

Odd nitrogen chemistry partially counteracts problems previously identified with calculated CO, Fig. 5. The shift, however, is quite small in comparison to the evident difference between model calculations and observations. The impact of odd nitrogen chemistry on a Venus model that uses more nominal chemical input data will be examined in a subsequent study.

The altitude-resolved assessment provided here indicates odd nitrogen chemistry needs to be considered in greater detail than the column-integrated results presented in the discovery paper [4]. In addition, neither the present nor any past study of odd nitrogen chemistry at these altitudes on Venus has examined the sensitivity of the model results to the details of the input chemistry data. There are substantial disagreements and/or uncertainties in the most

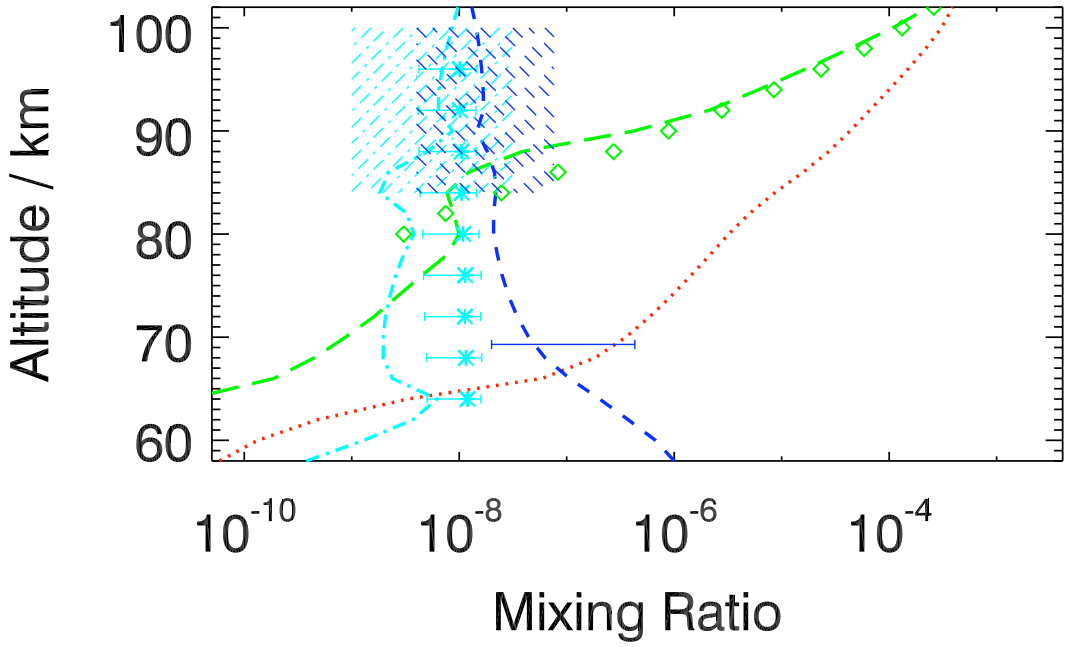


Fig. 3: Species abundances from model calculations with no odd nitrogen and selected observations. Long-dashed green line is modelled O, and green diamonds are inferred from 2006–7 observations [10]. Dotted red line is modelled O_2 for which there are no observed abundances. Short-dashed blue line is modelled SO_2; the solid blue range marker shows the range of values determined from rocket and space telescope observations over 1979–1995 [7]; and the short-dash hatched blue region shows the range of values inferred from sub-mm observations over 2004–8 [67]. Dash-dot cyan line is modelled SO; the cyan asterisks mark the profile inferred from 1991 space telescope observations [68]; and the dash-dot hatched cyan region shows the range of values inferred from sub-mm observations over 2007–8 [67]. The best-fit SO and SO_2 abundances from the sub-mm observations at $\sim 70 - 84$ km is zero and some observations show no evidence of SO and/or SO_2 at 84–100 km altitude [67].

important reactions that have not been resolved. This will be examined in a future study, along with temperature dependent cross sections for odd nitrogen species.

Table 2: Global-scale results and comparisons to observations.

Parameter	Observation	Model Calculation			
		0‡	5.5‡	15‡	30‡
O_2 column (10^{18} cm^{-2})	$< 0.8^a$, $< 1.5^b$	1.9	1.9	1.9	1.8
S_8 production (10^{10} cm^{-2} s^{-1})		1.7	1.7	1.6	1.5
H_2SO_4 production (10^{11}cm^{-2}s^{-1})	$\sim 2 - 10^c$	2.9	2.8	2.7	2.7

‡ NO mixing ratio (ppb) at 58 km lower boundary of model. [a] [16]. [b] [14]. [c] [3] and [75].

Conclusions

The effects of odd nitrogen chemistry in a Venusian mesospheric photochemical model have been examined for a model scenario in which heterogeneous oxidation of CO is important. There are potentially significant impacts on specific aspects of the modelled chemistry, e.g., SO and SO_2 abundances at 80–90 km, but the overall effects from the observed NO abundance are small. Further work is required to determine whether there are scenarios and specific aspects

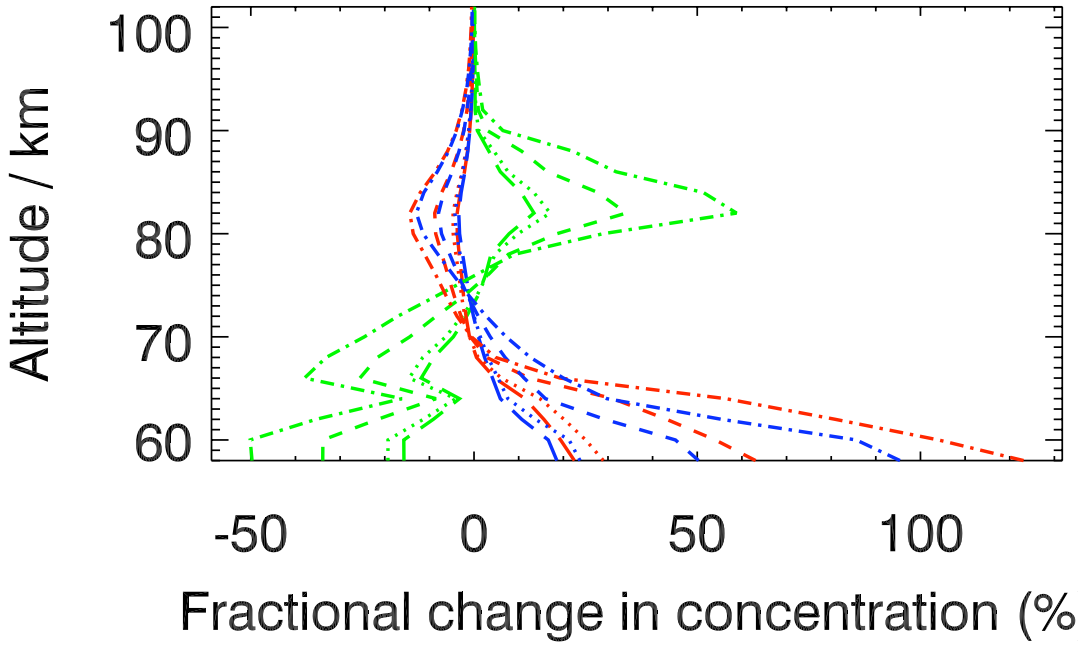


Fig. 4: Fractional difference in O (green), O_2 (red), and CO (blue) for a range of lower boundary conditions – NO(58 km) = 5.5 ppb (long-dash) as reported from observations [4], NO(58 km) = 7.0 ppb (dotted), NO(58 km) = 15 ppb (short-dash), and NO(58 km) = 30 ppb (dash-dot). The reference calculation has no NO_x but otherwise has identical input.

of Venus atmospheric chemistry that could be affected more substantially by odd nitrogen chemistry. Continued observational measurements of NO on Venus are required to establish the typical abundance of NO on Venus and its geographic and vertical distribution.

Acknowledgments

Partial funding for this research was provided by the Australian Research Council. Part of this work was performed at the Jet Propulsion Laboratory, California Institute of Technology, under contract to the National Aeronautics and Space Administration. Helpful comments from 2 reviewers are gratefully acknowledged.

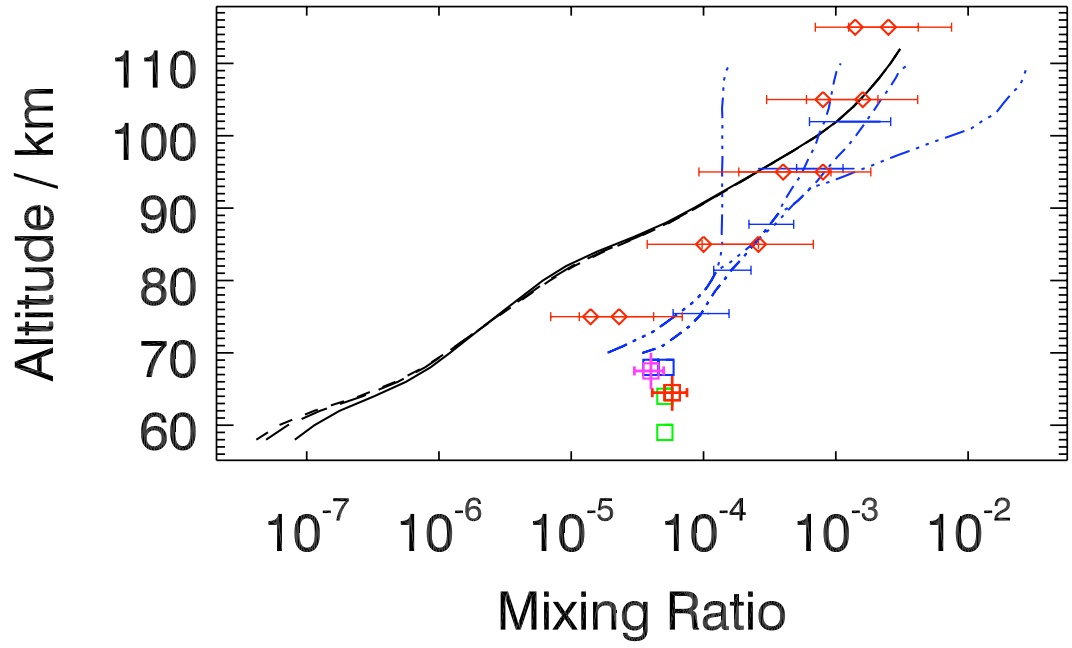


Fig. 5: Observed and modeled CO mixing ratios. Calculated profiles are for NO(58 km) = 0 (short-dash black), NO(58 km) = 5.5 ppb (long-dash black) as reported from observations [4], and NO(58 km) = 30 ppb (solid black). The dash-dot blue lines are the CO mixing ratios deduced from millimeter-wave observations at 6 am and 12 am Venus Local Time (VLT) in 1990 [69]. The dash-triple-dot blue lines are the CO mixing ratios deduced from millimeter-wave observations at 3 am and 9 am VLT in 1986 [69]. The green squares indicate the revised [57] day side CO mixing ratio derived from observations in 1966 [70]. The two altitude locations mark the likely altitude range for the observation based on the revised interpretation [57]. The magenta and red squares indicate the day side CO mixing ratios determined in 2006 and 2007, respectively [71], [72], and the blue squares indicate the morning and afternoon CO mixing ratios determined in 2007 and 2009 [73]. The red diamonds are the minimum and maximum CO mixing ratios observed on the night side in 1991 [74].

References

[1] F. P. Mills, M. Sundaram, T. G. Slanger, M. Allen, and Y. L. Yung, "Oxygen chemistry in the Venus middle atmosphere," in *Advances in Geosciences Volume 3: Planetary Science (PS)*, W.-H. Ip and A. Bhardwaj, Eds. Singapore: World Scientific Publishing Co., 2006, pp. 109–117.

[2] F. P. Mills and M. Allen, "A review of selected issues concerning the chemistry in Venus' middle atmosphere," *Plan. Sp. Sci.*, vol. 55, pp. 1729–1740, 2007.

[3] Y. Yung and W. DeMore, "Photochemistry of the stratosphere of Venus: Implications for atmospheric evolution," *Icarus*, vol. 51, pp. 199–247, 1982.

[4] V. A. Krasnopolsky, "A sensitive search for nitric oxide in the lower atmospheres of Venus and Mars: Detection on Venus and upper limit for Mars," *Icarus*, vol. 182, pp. 80–91, 2006.

[5] U. von Zahn, S. Kumar, H. Niemann, and R. Prinn, "Composition of the Venus atmosphere," in *Venus*, D. M. Hunten, L. Colin, T. M. Donohue, and V. I. Moroz, Eds. Tucson: Univ. of Arizona Press, 1983, pp. 299–430.

[6] L. W. Esposito, J.-L. Bertaux, V. Krasnopolsky, V. I. Moroz, and L. V. Zasova, "Chemistry of lower atmosphere and clouds," in *Venus II*, S. W. Bougher, D. M. Hunten, and R. J. Phillips, Eds. Tucson: Univ. of Ariz. Press, 1997, pp. 415–458.

[7] F. P. Mills, L. W. Esposito, and Y. L. Yung, "Atmospheric composition, chemistry, and clouds," in *Exploring Venus as a Terrestrial Planet*, L. W. Esposito, E. R. Stofan, and T. E. Cravens, Eds. Washington, DC: American Geophysical Union, 2007, ch. 5, pp. 73–100.

[8] P. Connes, J. Connes, F. Noxon, W. Traub, and N. Carlton, "$O_2(^1\Delta)$ emission in the day & night airglow of Venus," *Ap. J.*, vol. 233, pp. L29–L32, 1979.

[9] J. Bailey, V. S. Meadows, S. Chamberlain, and D. Crisp, "The temperature of the Venus mesosphere from $O_2(a^1\Delta_g)$ airglow observations," *Icarus*, vol. 197, pp. 247–259, 2008.

[10] J.-C. Gérard, A. Saglam, G. Piccioni, P. Drossart, C. Cox, S. Erard, R. Hueso, and A. Sánchez-Lavega, "Distribution of the O_2 infrared nightglow observed with VIRTIS on board Venus Express," *Geophys. Res. Lett.*, vol. 35, p. L02207, 2008.

[11] H. Nair, M. Allen, A. D. Anbar, Y. L. Yung, and R. T. Clancy, "A photochemical model of the Martian atmosphere," *Icarus*, vol. 111, pp. 124–150, 1994.

[12] F. P. Mills, "I. observations and photochemical modeling of the Venus middle atmosphere. II. thermal infrared spectroscopy of Europa and Callisto," Ph.D. Dissertation, California Institute of Technology, Pasadena CA, 1998, 366 pp.

[13] W. A. Traub and N. P. Carleton, "Observation of O_2, H_2O, and HD in planetary atmospheres," in *Exploration of the Planetary System*, A. Woszczyk and C. Iwaniszewska, Eds. Dordrecht, The Netherlands: Reidel, 1974, pp. 223–228.

[14] J. Trauger and J. Lunine, "Spectroscopy of molecular oxygen in the atmospheres of Venus and Mars," *Icarus*, vol. 55, pp. 272–281, 1983.

[15] F. P. Mills, "A spectroscopic search for molecular oxygen in the Venus middle atmosphere," *J. Geophys. Res.*, vol. 104, pp. 30 757–30 764, 1999.

[16] V. A. Krasnopolsky, "Chemical composition of Venus atmosphere and clouds: Some unsolved problems," *Plan. Sp. Sci.*, vol. 54, pp. 1352–1359, 2006.

[17] M. Shunmuga Sundaram, "Modeling the Venus Middle Atmosphere," MPhil Thesis, The Australian National University, Canberra, ACT, 2008.

[18] S. Solomon, R. R. Garcia, F. S. Rowland, and D. J. Wuebbles, "On the depletion of Antarctic ozone," *Nature*, vol. 321, pp. 755–758, 1986.

[19] C. Mills and L. Phillips, "Photo-oxidation of CO in a sulphuric acid aerosol," *J. Photochem. Photobiol. A: Chem.*, vol. 74, pp. 7–9, 1993.

[20] C. Mills, G. Rowland, J. Westergren, and L. Phillips, "Quantum yields of co_2 and so_2 formation from 193 nm photo-oxidation of co in a sulfuric acid aerosol," *J. Photochem. Photobiol. A: Chem.*, vol. 93, pp. 83–87, 1996.

[21] S. Wrenn, L. Butler, G. Rowland, C. Knox, and L. Phillips, "The necessity for multiphoton processes in the 193-nm photochemistry of sulphuric acid aerosols," *J. Photochem. Photobiol. A: Chem.*, vol. 129, pp. 101–104, 1999.

[22] A. I. F. Stewart, J.-C. Gerard, D. W. Rusch, and S. W. Bougher, "Morphology of the Venus ultraviolet night airglow," *J. Geophys. Res.*, vol. 85, pp. 7861–7870, 1980.

[23] Y. L. Yung and M. B. McElroy, "Fixation of nitrogen in the prebiotic atmosphere," *Science*, vol. 203, pp. 1002–1004, 1979.

[24] H. Pernice, P. Garcia, H. Willner, J. S. Francisco, F. P. Mills, M. Allen, and Y. L. Yung, "Laboratory evidence for a key intermediate in the Venus atmosphere: Peroxychloroformyl radical," *PNAS*, vol. 101, pp. 14 007–14 010, 2004.

[25] D. M. Hunten, "Venus lightning: Pros and cons," *Adv. Sp. Res.*, vol. 15, pp. 109–112, 1995.

[26] C. T. Russell, T. L. Zhang, M. Delva, W. Magnes, R. J. Strangeway, and H. Wei, "Lightning on Venus inferred from whistler-mode waves in the ionosphere," *Nature*, vol. 450, pp. 661–662, 2007.

[27] P. Thamm, "Research on venus's atmospheric chemistry," 2010, undergraduate student project report.

[28] F. P. Mills, "Water vapour in the Venus middle atmosphere," *Advances in Space Research*, vol. 23, pp. 1573–1576, 1999.

[29] M. Allen, Y. L. Yung, and J. W. Waters, "Vertical transport and photochemistry in the terrestrial mesosphere and lower thermospehere," *Journal of Geophysical Research*, vol. 86, pp. 3617–3627, 1981.

[30] A. Seiff, "Models of Venus's atmospheric structure," in *Venus*, D. M. Hunten, L. Colin, T. M. Donohue, and V. I. Moroz, Eds. Tucson: Univ. of Arizona Press, 1983, pp. 1045–1048.

[31] D. Crisp, "Radiative forcing of the Venus mesosphere I: Solar fluxes and heating rates," *Icarus*, vol. 67, pp. 484–514, 1986.
[32] W. B. DeMore, S. P. Sander, D. M. Golden, R. F. Hampson, M. J. Kurylo, C. J. Howard, A. R. Ravishankara, C. E. Kolb, and M. J. Molina, "Chemical kinetics and photochemical data for use in stratospheric modeling: Evaluation number 12," Jet Propulsion Laboratory, Pasadena CA, JPL Publication 97-4, 1997.
[33] S. P. Sander *et al*., "Chemical kinetics and photochemical data for use in stratospheric modeling evaluation number 14," Jet Propulsion Laboratory, Pasadena CA, JPL Publication 02-25, 2002, http://jpldataeval.jpl.nasa.gov/.
[34] J. Nicovich, K. Kreutter, and P. Wine, "Kinetics and thermochemistry of ClCO formation from the Cl+CO association reaction," *J. Chem. Phys*., vol. 92, pp. 3539–3544, 1990.
[35] B. R. Lewis and J. H. Carver, "Temperature dependence of the carbon dioxide photoabsorption cross section between 1200 and 1970 Angstrom," *J. Quant. Spectrosc. Radiat. Transfer*, vol. 30, pp. 297–309, 1983.
[36] D. E. Shemansky, "CO_2 extinction coefficient 1700 - 3000 Angstrom," *J. Chem. Phys*., vol. 56, pp. 1582–1587, 1972.
[37] D. E. Freeman, K. Yoshino, J. R. Esmond, and W. H. Parkinson, "High resolution absorption cross-section measurements of SO_2 at 213 k in the wavelength region 172 - 240 nm," *Planet. Space Sci*., vol. 32, pp. 239–248, 1984.
[38] C. H. Hearn and J. A. Joens, "The near uv absorption spectrum of CS_2 and SO_2 at 300 K," *J. Quant. Spectrosc. Radiat. Transfer*, vol. 45, pp. 69–75, 1991.
[39] B. Leroy, P. Rigaud, J. L. Jourdain, and G. LeBras, "Spectres d-absorption dans le proche ultraviolet de CS_2 et SO_2 entre 200 et 300 K," *The Moon and the Planets*, vol. 29, pp. 177–183, 1983.
[40] S. L. Manatt and A. L. Lane, "A compilation of the absorption cross-sections of SO_2 from 106 to 403 nm," *J. Quant. Spectrosc. Radiat. Transfer*, vol. 50, pp. 267–276, 1993.
[41] R. D. Martinez and J. A. Joens, "SO_2 absorption cross-section measurements from 197 nm to 240 nm," *Geophys. Res. Lett*., vol. 19, pp. 277–279, 1992.
[42] T. J. McGee and J. J. R. Burris, "SO_2 absorption cross sections in the near UV," *J. Quant. Spectrosc. Radiat. Transfer*, vol. 37, pp. 165–182, 1987.
[43] K. E. Sprague and J. A. Joens, "SO_2 absorption cross-section measurements from 320 nm to 405 nm," *J. Quant. Spectrosc. Radiat. Transfer*, vol. 53, pp. 349–352, 1995.
[44] M. Suto, R. L. Day, and L. C. Lee, "Fluorescence yields from photodissociation of SO_2 at 1060 - 1330 Angstrom," *J. Phys. B.: At. Mol. Phys*., vol. 15, pp. 4165–4174, 1982.
[45] S. M. Anderson and K. Mauersberger, "Laser measurements of ozone absorption cross sections in the Chappuis Band," *Geophys. Res. Lett*., vol. 19, pp. 933–936, 1992.
[46] J. B. Burkholder and R. K. Talukdar, "Temperature dependence of the ozone absorption spectrum over the wavelength range 410 to 760 nm," *Geophys. Res. Lett*., vol. 21, pp. 581–584, 1994.
[47] D. E. Freeman, K. Yoshino, J. R. Esmond, and W. H. Parkinson, "High resolution absorption cross-section measurements of ozone at 195 k in the wavelength region 240 - 350 nm," *Planet. Space Sci*., vol. 32, pp. 239–248, 1984.
[48] M. Griggs, "Absorption coefficients of ozone in the ultraviolet and visible regions," *J. Chem. Phys*., vol. 49, pp. 857–859, 1968.
[49] L. T. Molina and J. J. Molina, "Absolute absorption cross sections of ozone in the 185- to 350-nm wavelength range," *J. Geophys. Res*., vol. 91, pp. 14 501–14 508, 1986.
[50] Y. Tanaka, E. C. Y. Inn, and K. Watanabe, "Absorption coefficients of gases in the vacuum ultraviolet, part IV. ozone," *J. Chem. Phys*., vol. 21, pp. 1651–1653, 1953.
[51] K. Yoshino, D. W. Freeman, J. R. Esmond, and W. H. Parkinson, "Absolute absorption cross-section measurements of ozone in the wavelength region 238-335 nm and the temperature dependence," *Planet. Space Sci*., vol. 36, pp. 395–398, 1988.
[52] K. Yoshino, J. R. Esmond, D. E. Freeman, and W. H. Parkinson, "Measurements of absolute absorption cross sections of ozone in the 185- to 254-wavelength region and the temperature dependence," *J. Geophys. Res*., vol. 98, pp. 5205–5211, 1993.
[53] L. T. Molina, J. J. Lamb, and M. J. Molina, "Temperature dependent UV absorption cross sections for carbonyl sulfide," *Geophys. Res. Lett*., vol. 8, pp. 1008–1011, 1981.
[54] W. A. Lane and R. Opstbaum, "High altitude Venus haze from Pioneer Venus limb scans," *Icarus*, vol. 54, pp. 48–58, 1983.
[55] U. von Zahn, K. H. Fricke, D. M. Hunten, D. Krankowsky, K. Mauersberger, and A. O. Nier, "The upper atmosphere of Venus during morning conditions," *J. Geophys. Res*., vol. 85, pp. 7829–7840, 1980.
[56] R. Woo and A. Ishimaru, "Eddy diffusion coefficient for the atmosphere of Venus from radio scintillation measurements," *Nature*, vol. 289, pp. 383–384, 1981.
[57] L. D. G. Young, "High resolution spectrum of Venus–a review," *Icarus*, vol. 17, pp. 632–658, 1972.
[58] P. Connes, J. Connes, L. Kaplan, and W. Benedict, "Traces of HCl and HF in the atmosphere of Venus," *Ap. J*., vol. 147, pp. 1230–1237, 1967.
[59] B. Bezard, C. deBergh, D. Crisp, and J.-P. Maillard, "The deep atmosphere of Venus revealed by high-resolution nightside spectra," *Nature*, vol. 345, pp. 508–511, 1990.
[60] S. P. Sander *et al*., "Chemical kinetics and photochemical data for use in atmospheric studies evaluation number 15," Jet Propulsion Laboratory, Pasadena CA, JPL Publication 06-2, 2006, http://jpldataeval.jpl.nasa.gov/.
[61] I. M. Campbell and B. A. Thrush, "Behaviour of carbon dioxide and nitrous oxide in active nitrogen," *Trans. Faraday Soc*., vol. 62, pp. 3366–3374, 1966.
[62] ——, "The association of oxygen atoms and their combination with nitrogen atoms," *Proc. Roy. Soc. London A Math. Phys. Sci*., vol. 296, pp. 222–232, 1967.

[63] L. F. Phillips and H. I. Schiff, "Mass spectrometric studies of atom reactions. i. reactions in the atomic nitrogen-ozone system," *J. Chem. Phys.*, vol. 36, pp. 1509–1517, 1962.
[64] J. L. Fox, "Rate coefficient for the reaction N + NO," *J. Geophys. Res.*, vol. 99, pp. 6273–6276, 1994.
[65] R. F. Hampson, "Chemical kinetic and photochemical data sheets for atmospheric reactions," Federal Aviation Administration, Washington, DC, Federal Aviation Administration Report FAA-EE-80-17, 1980.
[66] D. L. Baulch, R. A. Cox, J. R. F. Hampson, J. A. Kerr, J. Troe, and R. T. Watson., "Evaluated kinetic and photochemical data for atmospheric chemistry,," *J. Phys. Chem. Ref. Data*, vol. 9, pp. 295–471, 1980.
[67] B. J. Sandor, R. T. Clancy, G. Moriarty-Schieven, and F. P. Mills, "Sulfur chemistry in the Venus mesosphere from SO_2 and SO microwave spectra," *Icarus*, vol. 208, pp. 49–60, 2010.
[68] C. Y. Na, L. W. Esposito, W. E. McClintock, and C. A. Barth, "Sulfur dioxide in the atmosphere of Venus," *Icarus*, vol. 112, pp. 389–395, 1994.
[69] R. Clancy and D. Muhleman, "Long-term (1979-1990) changes in the thermal, dynamical, and compositional structure of the Venus mesosphere as inferred from microwave spectral line observations of ^{12}CO, ^{13}CO, and $C^{18}O$," *Icarus*, vol. 89, pp. 129–146, 1991.
[70] P. Connes, J. Connes, L. D. Kaplan, and W. S. Benedict, "Carbon monoxide in the Venus atmosphere," *Ap. J.*, vol. 152, pp. 731–743, 1968.
[71] P. G. J. Irwin, R. de Kok, A. Negrão, C. C. C. Tsang, C. F. Wilson, P. Drossart, G. Piccioni, D. Grassi, and F. W. Taylor, "Spatial variability of carbon monoxide in venus' mesosphere from Venus Express/Visible and Infrared Thermal Imaging Spectrometer measurements," *J. Geophys. Res.*, vol. 113, p. E00B01, 2008.
[72] N. Iwagami, T. Yamaji, S. Ohtsuki, and G. L. Hashimoto, "Hemispherical distribution of CO above the Venus' clouds by ground-based 2.3 μm spectroscopy," *Icarus*, vol. 207, pp. 558–563, 2010.
[73] V. A. Krasnopolsky, "Spatially-resolved high-resolution spectroscopy of Venus 1. Variations of CO_2, CO, HF, and HCl at the cloud tops," *Icarus*, vol. 208, pp. 539–547, 2010.
[74] E. Lellouch, J. J. Goldstein, J. Rosenqvis, S. W. Bougher, and G. Paubert, "Global circulation, thermal structure, and carbon monoxide distribution in Venus' mesosphere in 1991," *Icarus*, vol. 110, pp. 315–339, 1994.
[75] J. R. Winick and A. I. Stewart, "Photochemistry of SO_2 in Venus' upper cloud layers," *J. Geophys. Res.*, vol. 85, pp. 7849–7860, 1980.

Eucrite Meteorites: Clues to Early Igneous Processes on Differentiated Asteroids

Marc D. Norman[1], Katherine Bermingham[1]*, Andrew G. Christy[1**], Vickie Bennett[1]

[1]*Research School of Earth Sciences, Australian National University, Canberra ACT 0200 Australia*

**current address: Institut für Mineralogie, Westfälische Wilhelms-Universität, Münster 48149 Germany*
***current address: Centre for Advanced Microscopy, Australian National University, Canberra, ACT 0200 Australia*

Summary: Eucrite meteorites are among the oldest basalts in the Solar System. They are thought to sample the primary crust of asteroid 4Vesta based on remote sensing spectroscopy and plausible mechanisms for delivery of the meteorites to Earth. This project aims to better understand processes of early planetary differentiation through study of the petrology and geochemistry of a suite of basaltic eucrites. Textures of these eucrites indicate emplacement depths ranging from the surface of the asteroid to a few km. Major element characteristics of all samples are similar and consistent with melt compositions, whereas trace elements are highly variable. Terrestrial weathering has altered the trace element and radiogenic isotopic compositions of some samples, but primary features reflecting either partial melting of early formed cumulates or extraction of melts from basaltic crust are preserved. The early evolution of the eucrite parent body was more complex than accounted for by current models.

Keywords: eucrite, meteorite, asteroid, 4Vesta.

Introduction

Eucrites are extraterrestrial basalts that form part of the HED (howardite, eucrite and diogenite) achondrite group of meteorites. Understanding the petrogenesis of these basalts is important for evaluating mechanisms of chemical and physical differentiation on small planesesimals during the early history of the Solar System. The aim of our research was to evaluate the petrology and geochemistry of a suite of eucrite meteorites that were collected in Antarctica in an effort to identify the processes which produced these magmas and their source regions.

The HED parent body, thought to be asteroid 4Vesta, is unique among meteorite parent bodies in that it provides an unusual example of a highly differentiated planetesimal with a well preserved, primary basaltic crust [1]. By understanding its early chemical differentiation, planetary scientists hope to comprehend the origin and early evolution of larger bodies such as the Earth, Moon, and Mars. Study of eucrites also provides valuable groundtruth for the current *Dawn* spacecraft mission that will investigate two of the largest main-belt asteroids, Vesta and Ceres, which are endmember examples of the formation and evolution of proto-planets [2, 3].

Eucrites possess magmatic textures, mineral compositions and trace element patterns that indicate an igneous origin [4, 5]. Most eucrites have been brecciated and metamorphosed by impact events and prolonged heating, which complicates determination of their primary igneous evolution. For this study we were allocated thin sections and/or whole rock samples of 13 monomict, mostly unbrecciated eucrites from the NASA Antarctic meteorite collection (Table 1). We requested these samples because they are among the least modified by extraterrestrial or terrestrial alteration, and because they are among the least brecciated examples of eucrites in the Antarctic meteorite collection. In this paper we first present brief petrographic descriptions of these samples based on

observations of the thin sections allocated to us (Table 1), supplemented by information available in the literature. We then present major and trace element data that suggest the discovery of a new class of basaltic eucrite, and conclude with a brief discussion of terrestrial weathering effects on eucrite geochemistry and the origins of eucrite magmas.

Table 1. *Summary characteristics of Antarctica eucrite samples in this study. Data from the Meteoritical Bulletin database (http://tin.er.usgs.gov/meteor/about.php) and the NASA Antarctic meteorite curatorial database (http://curator.jsc.nasa.gov/antmet/index.cfm).*

Sample Number	Classification	Weathering (A=minor rust, B=moderate rust)	Collection locality Original Mass	Year of collection	Allocation, this study TS=thin section, WR=whole rock (mass)
BTN 00300	brecciated monomict	A	Bates Nunataks 124.56 g	2000	BTN00309,27 (TS) BTN00309,28 (TS) BTN00309,23 (WR, 3.8 g)
EET 90020	unbrecciated monomict	A	Elephant Moraine 154.02 g	1990	EET90020,42 (TS) EET90020,44 (TS) EET90020,38 (WR, 4.3 g)
EET 90029	unbrecciated monomict	A/B	Elephant Moraine 7.920 g	1990	EET90029,7 (WR, 0.5 g)
GRA 98098	brecciated monomict	B	Graves Nunataks 779.2 g	1998	GRA98098,49 (TS) GRA98098,50 (TS) GRA98098,45 (WR, 3.7 g)
MAC 02522	unbrecciated monomict	B	MacAlpine Hills 5.70 g	2002	MAC02522,14 (WR 0.6 g)
MAC 041169	unbrecciated monomict	B	MacAlpine Hills 3.19 g	2004	MAC041169,4 (TS) MAC041169,5 (WR 0.1 g)
MET01081	unbrecciated monomict	A/B	Meteorite Hills 27.40 g	2001	MET01081,15 (WR 0.5 g)
PCA 82502	brecciated monomict	A	Pecora Escarpment 890.40 g	1982	PCA82502,45 (TS) PCA82502,94 (TS) PCA82502,89 (WR, 3.9 g)
PCA 91078	unbrecciated monomict	A/B	Pecora Escarpment 20.85 g	1991	PCA91078,18 (WR, 0.6 g)
QUE 94484	unbrecciated monomict	B	Queen Alexandra Range 5.56 g	1994	QUE94484,4 (WR, 0.5 g)
QUE 97014	unbrecciated monomict	B	Queen Alexandra Range 142.30 g	1997	QUE97014,10 (TS) QUE97014,26 (TS) QUE97014,22 (WR, 3.9 g)
QUE 99658	unbrecciated monomict	A	Queen Alexandra Range 8.77 g	1999	QUE99658,6 (WR 0.5 g)
RKP 80224	unbrecciated monomict	A/B	Reckling Peak 8.01 g	1980	RKP80224,9 (WR 0.5 g)

Petrography

All of the eucrites studied here are basaltic with textures indicating crystallization of a silicate melt, followed by variable degrees of shock, brecciation, and thermal metamorphism likely related to later impacts that reworked the shallow igneous crust of the parent asteroid. The mineralogy of these eucrites consists primarily of calcic plagioclase ($Na[AlSi_3O_8]$ to $Ca[Al_2Si_2O_8]$) and pyroxene ($(Fe,Mg,Ca)_2Si_2O_6$). Minor and accessory phases include silica polymorphs (SiO_2, as either quartz or tridymite), chromite, ilmenite, troilite, olivine, K-feldspar, native iron, and rare phases enriched in incompatible trace elements such as Zr and REE [4, 5, 6].

Textures can be classified primarily as ophitic to subophitic, with grain sizes ranging from relatively coarse-grained varieties such as GRA 98098 and EET 90020 to finer-grained examples such as PCA 82502 (Fig. 1). The combination of well-formed, euhedral plagioclase crystals, with the sharp, straight grain contacts between plagioclase and pyroxene, the equant grain size, and the presence of exsolution lamellae in pyroxenes in GRA 98098 and EET 90020, and melt inclusions

in the cores of many plagioclase grains such as those in EET 90020 (Fig. 1) provide clear evidence for igneous crystallization and relatively slow subsolidus cooling. In contrast, the crenulated, granular textures of BTN 00300 and QUE 97014 characterised by anhedral to subhedral laths of plagioclase, irregular grain size and shapes, and the presence of curved grain boundaries (Fig. 1) suggest the igneous precursors were modified by post-crystallisation brecciation and metamorphism, probably by early impacts while the parent asteroid was still hot enough to drive recrystallisation. Recrystallised areas also occur in GRA 98098 and EET 90020, indicating a multi-stage history for these samples as well. The presence of fractures, undulose extinction in the plagioclase, and clouding or mosaicism of pyroxenes such as that in observed MAC 041169 (Fig. 1) provide evidence for shock produced by later impacts. Most of the eucrites studied here have glassy rinds or fusion crusts coating their exterior surfaces (e.g. PCA 82502; Fig. 1); these were produced by melting during passage of the meteorite through the Earth's atmosphere. Aqueous alteration produced during residence on the Earth is recognized by the patches of Fe-oxides ("rust") such as that visible on the left margin of the image of MAC 041169 (Fig. 1) . Additional details of the petrography of these and other unbrecciated basaltic eucrites are given by [6, 7, 8, 9].

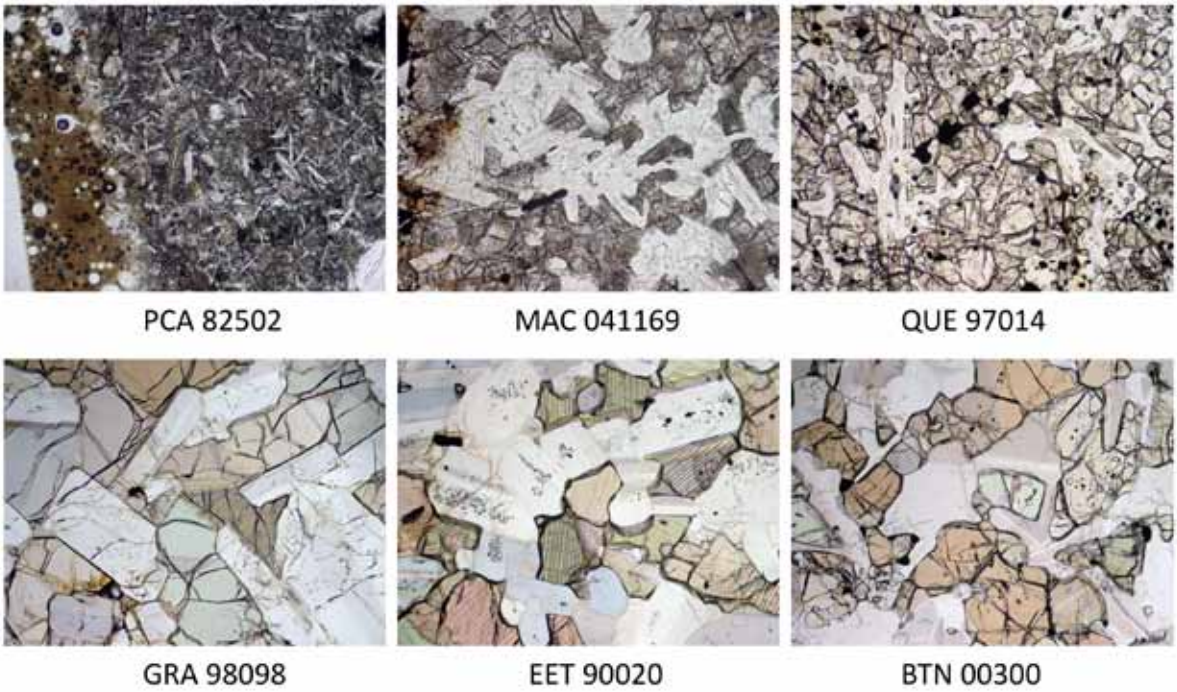


Figure 1a. *Optical photomicrographs of eucrite thin sections illustrating the range of textures and grain sizes of samples in this study. Field of view for all images is 1 mm wide.*

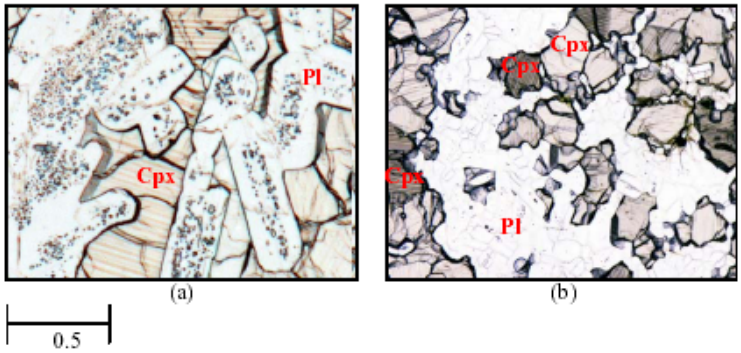


Figure 1b*. Photomicrographs of 90020 comparing domains with igneous (left) and metamorphic (right) textures. Igneous domains have lath-shaped plagioclase (Pl) with abundant melt inclusions, and pyroxene (Cpx) with exsolution lamellae****.*** *Metamorphic domains are characterized by clear, equant grains of plagioclase and a lack of exsolution lamellae in the Cpx.*

Geochemistry

Methods

Whole rock chips were crushed to sand-sized fragments using a steel piston and a hammer, and ground by hand into a fine powder using an agate mortar and pestle. Chips with adhering fusion crust were removed and set aside for later analysis. Aliquots of the whole rock powders were fused to a glass with a lithium borate flux for major and trace element analysis. Conventional disks for major element analysis by X-ray fluorescence (XRF) were prepared from five of the samples using 0.27 g of sample and 1.72 g of flux. In addition, small glass beads, or 'lentils' were prepared from all samples using 0.1 g of sample and 0.3 g of flux. The major elements Na, Mg, Al, Si, P, K, Ca, Ti, Mn, and Fe were measured using a Phillips (PANalytical) PW2400 wavelength-dispersive XRF spectrometer. Data are reported as oxides with FeO* = total Fe reported as FeO (Table 2). Major element analyses were also obtained on the lentil beads mounted in polished epoxy disks using a JEOL JSM6400 scanning electron microscope equipped with an energy-dispersive detector and Link ISIS quantification software. The XRF analyses were calibrated against a suite of 28 international standard rock powders. The EDS-SEM analyses were calibrated initially against individual mineral standards, and the data adjusted by comparison with the XRF data for corresponding samples. The SEM vs. XRF adjustment factors were ≤2% relative for SiO_2, Al_2O_3, FeO* and MgO, 4% for CaO, and 40% for TiO_2 and Na_2O. The XRF and SEM data were averaged and presented in Table 2.

Trace element compositions were measured by laser ablation ICPMS analysis of the glass lentils using an Agilent HP7500 quadrupole ICPMS coupled to a pulsed excimer UV (193 nm) laser. The samples were ablated by traversing polished surfaces of the glasses under the laser beam across at a speed of 0.5 mm/minute. Ablation was conducted in a custom-built, low-volume sample cell under a He + H atmosphere using laser spot diameter of 86 μm and a laser pulse rate of 10 Hz. Ablated material was mixed with Ar carrier gas prior to introduction into the mass spectrometer. Each analysis consisted of 120 seconds of data collection in time resolved mode, including 30 seconds on the dry gas background and 90 seconds of ablation on the sample. Element sensitivity was calibrated against the NIST 612 glass and each analysis was normalized to the CaO wt% content of the sample to correct for variations in ablation yield. Following reduction of the data to concentrations, the analyses were adjusted to preferred values of the basaltic reference material BHVO-2 prepared and run with the samples to correct for second-order matrix effects relative to the NIST 612 glass calibration.

^{147}Sm-^{143}Nd isotopic compositions of the eucrites were measured by thermal ionization mass spectrometry. Powdered samples were spiked with a mixed solution containing ^{150}Nd and ^{147}Sm for isotope dilution measurements, and dissolved using distilled HF and HCl. The Sm and Nd were isolated from matrix elements by cation exchange chromatography using Biorad AG 50W-X8 and HDEHP resins. Sm and Nd isotopic compositions were measured as metals on a Finnigan MAT 261 thermal ionization mass spectrometer using Ta-Re double filament assemblies. Isotope ratios were corrected for mass fractionation by normalization to $^{146}Nd/^{144}Nd=0.7219$ and $^{149}Sm/^{152}Sm=0.516848$. Mean $^{143}Nd/^{144}Nd$ ratios of each sample were corrected to a value of 0.511860 for the La Jolla reference standard.

Results

Major elements

Basaltic eucrites are classified as cumulate and noncumulate varieties [4, 10, 11, 12]. Cumulate eucrites have major and trace element compositions reflecting accumulation of mineral phases in excess of those predicted for closed-system crystallization of a silicate melt. Noncumulate eucrites

have compositions matching those predicted for a magma formed by either melting or fractional crystallization without any excess crystals. Cumulate eucrites have higher Mg/Fe, lower Ca/Al, and lower contents of incompatible elements (i.e. elements that partition preferentially into the melt phase) than noncumulate eucrites, reflecting enrichments in early-formed pyroxene and plagioclase relative to the silicate melt (Figure 2). Major element characteristics of the meteorites studied here identify all of them as noncumulate eucrites (Figure 2) with no distinction between the Group 1 and Group 2 samples that are identified by trace element compositions (next section). No effects of terrestrial weathering are apparent in the major element compositions.

Trace elements

Trace element compositions of the Antarctic eucrites studied here illustrate the geochemical diversity of basaltic eucrites and some of the current controversies concerning eucrite petrogenesis. Two compositional groups are recognised based on trace element characteristics. Group 1 eucrites have compositions typical of noncumulate eucrites. Chondrite-normalised rare-earth element (REE) patterns of the Group 1 eucrites are generally flat with a slight enrichment of light REE (LREE; La, Ce) relative to the heavy REE (HREE; Yb, Lu), and increasing depletion of Eu (negative Eu anomaly) relative to the LREE with increasing REE abundances (Fig. 3). Concentrations of other incompatible trace elements in these Group 1 eucrites are also well correlated and fall along the main trend of noncumulate basaltic eucrite compositions (Fig. 4).

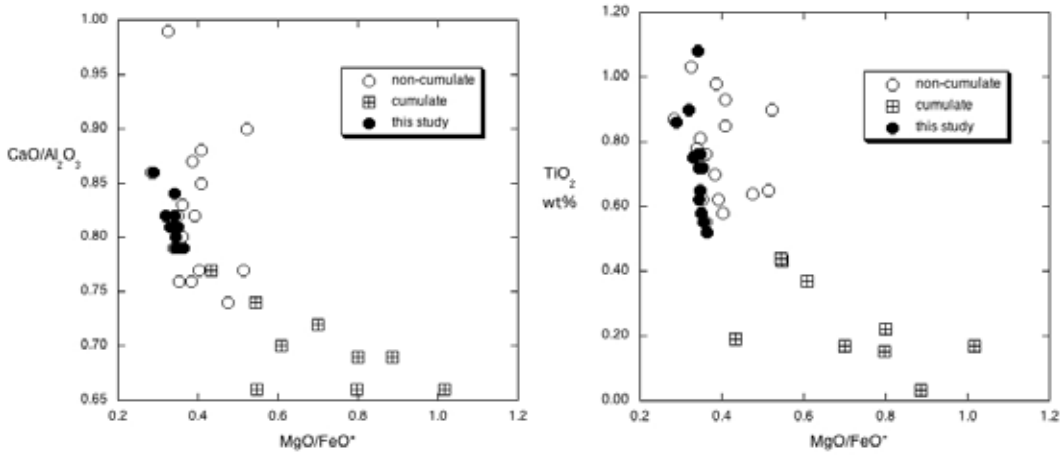


Figure 2. Major element compositions of the meteorites in this study identify them as noncumulate eucrites. Literature data from [10, 11].

Group 2 eucrites have lower concentrations of REE and other incompatible lithophile elements, with samples PCA 91078 and MET 01081 having among the lowest concentrations of these trace elements yet reported for noncumulate eucrites (Figs. 3, 4). Group 2 samples are distinguished by depletions of LREE relative to HREE, and positive Eu anomalies. Sample PCA 82502 has a positive Ce anomaly indicating terrestrial weathering as discussed in the next section. REE patterns of PCA 91078 and MET 01081 are similar to those of the cumulate eucrite Moore County (Fig. 3). Yamaguchi et al. [8] also described examples of LREE-depleted noncumulate eucrites.

Trace elements that partition into the major mineral phases of eucrites rather than the melt (e.g. Cr into pyroxene and spinel, Sr into plagioclase) show a limited range of concentrations and no systematic differences between cumulate and noncumulate eucrites (Fig. 5). Similar relationships are observed between our Group 1 and Group 2 samples. In addition, our data suggest a negative correlation between the abundance of Cr and incompatible elements such as Hf and the REE that links the Group 1, Group 2, and cumulate eucrites, whereas a similar trend is not evident in the literature data for noncumulate eucrites (Figs. 5, 6a).

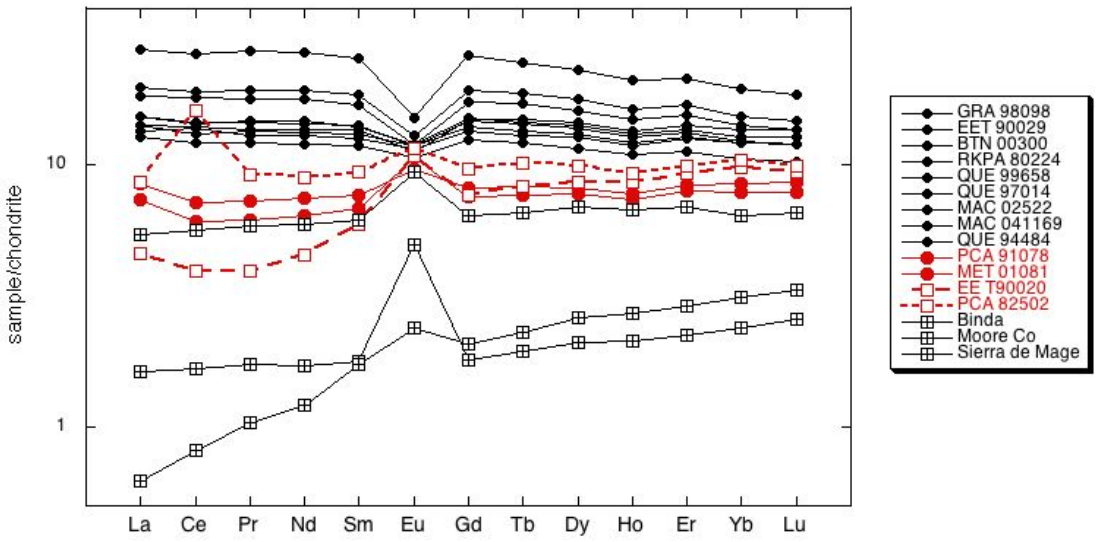


Figure 3. Chondrite-normalised REE patterns for Antarctic eucrites (this study) and representative cumulate eucrites (cross-squares; [10, 11]). Group 1 eucrites plotted as filled circles. Group 2 eucrites plotted as red symbols.

^{147}Sm-^{143}Nd isotopic compositions of the eucrites in this study span a broad range, reflecting the time-intergated evolution of their whole rock LREE patterns. Fig. 7 shows the ^{143}Nd/^{144}Nd isotopic compositions of the eucrites ploted against their ^{147}Sm/^{144}Nd ratios in a conventional isochron diagram. EET 90020 has the highest ^{143}Nd/^{144}Nd measured in this study, reflecting its extreme LREE depletion (Fig. 3). Most of the compositions fall along a 4.5 Ga reference line, except for PCA 82502, but even excluding this sample the data do not form an acceptable isochron indicating excess scatter. Table 4 shows the Sm-Nd isotopic data calculated as model ages relative to primitive chondritic reference compositions (CHUR; CHondritic Uniform Reservoir). In principle, CHUR model ages give an indication of the timing of Sm/Nd fractionation relative to a chondritic source composition. CHUR model ages of the Antarctic eucrites range from 0.6-0.8 Ga for MAC 02522 to unrealistically old values (greater than the age of the solar system) for EET 90029, RKPA 80224 and PCA 91078.

Discussion

Terrestrial weathering of Antarctic eucrites

The eucrites studied here were all collected as 'finds' in the Antarctic. A few Antarctic eucrites have relatively young terrestrial residence ages (e.g. 2500 years for ALH 85001) but most Antarctic eucrites have terrestrial residence ages older than the limit of conventional ^{14}C dating (≥40,000 years) [14], with several examples having residence ages of 100,000-300,000 years [15, 16]. Although weathering rates in the Antarctic deep freeze are very slow, this is sufficient time for secondary alteration phases such as salts, Fe-oxides, and clay minerals to form along cracks and grain boundaries [17]. Chemical effects such as positive and negative Ce anomalies (reflecting oxidized Ce), enrichment of Se, and mobility of alkali elements such as K and Rb have also been noted in Antarctic eucrites [13, 16, 17]. Mittlefehldt and Lindstrom [13, 17] included depletion of LREE and positive Eu anomalies as likely indicators of terrestrial weathering of basaltic eucrites, citing preferential dissolution of REE-enriched phosphate minerals as the likely mechanism with Eu retained in plagioclase. They further proposed that immobile minor and trace elements such as Ti, Zr, Hf, Nb, and Ta better preserve the primary geochemical characteristics of Antarctic eucrites because these elements are hosted by minerals more resistant to weathering (e.g. zircon, ilmenite).

Based on these criteria, the positive Ce anomaly of PCA 82502 (Fig. 3) is a clear indicator of terrestrial weathering. Depletion of U relative to Th in some samples (Fig. 6b) probably also reflects mobility of oxidized U. Displacement of the Sm-Nd isotopic composition of this meteorite to the right of the 4.5 Ga reference line (Fig. 7) is further evidence of recent LREE-depletion after a long-term evolution with a near-chondritic Sm/Nd. The Zr, Hf, Nb, Ta, Ba, and Th contents of PCA 82502 are similar to those of the Group 1 eucrites and well within the range for typical noncumulate eucrites (Fig. 4), whereas its Sm concentration appears to be ~20-25% low relative to its Hf content (Fig. 8), further supporting the inference that the LREE-depleted pattern of this meteorite is due to terrestrial weathering. This conclusion raises the question of whether LREE depletions in the other Group 2 eucrites are also due to weathering rather than reflecting primary igneous compositions. For these other Group 2 eucrites, however, the effects of terrestrial weathering are not so clear. Abundances of the immobile trace elements (Zr, Hf, Nb, Ta, Ba, Th) are at the low end of the range observed for noncumulate eucrites and transitional to compositions typically associated with cumulate eucrites (Figs. 4, 6; Table 3), and the low Sm contents of PCA 91078 and MET 01081 are as expected for primary igneous compositions by comparison with their Hf concentrations (Fig. 8). In addition, the 4.5-4.6 Ga CHUR model ages are consistent with long-term LREE depletion and no recent modification of the Sm/Nd of these meteorites. We conclude that the trace element compositions of PCA 91078 and MET 01081 reflect primary igneous characteristics, whereas PCA 82502 was modified by terrestrial weathering. The trace element composition of EET 90020 (Table 2, Fig. 3) may reflect partial melting or metamorphism on the eucrite parent body, as discussed in the following section [7, 8].

Table 2. Major element compositions of Antarctic eucrites analysed for this study. All data in wt% oxides.

	GRA 98098	EET90029	BTN 00300	RKPA 80224	QUE 99658	QUE 97014	MAC 02522	4AC 04116	QUE 94484	PCA 91078	MET 01081	EET 90020	PCA 82502
SiO_2	49.47	48.99	49.27	48.96	48.97	49.50	48.83	48.28	47.86	48.15	49.14	49.36	49.55
TiO_2	0.57	0.62	0.64	0.72	0.72	0.75	0.86	0.90	1.08	0.52	0.55	0.62	0.77
Al_2O_3	13.06	12.65	13.17	12.87	12.63	13.02	12.04	12.54	11.74	12.47	13.27	12.93	12.72
FeO*	18.96	18.32	18.66	18.61	18.89	18.53	20.52	18.98	18.22	19.75	18.59	18.77	18.90
MnO	0.56	0.54	0.56	0.53	0.54	0.55	0.61	0.53	0.49	0.59	0.55	0.57	0.56
MgO	6.63	6.35	6.47	6.40	6.68	6.11	5.94	6.06	6.23	7.20	6.59	6.46	6.47
CaO	10.58	10.29	10.40	10.22	10.02	10.54	10.35	10.30	9.88	9.92	10.45	10.30	10.40
Na_2O	0.28	0.49	0.49	0.55	0.58	0.45	0.69	0.57	0.59	0.34	0.36	0.53	0.45
K_2O	0.05	---	0.03	---	---	0.06	---	---	---	---	---	0.04	0.04
P_2O_5	0.10	0.06	0.06	0.04	0.06	0.07	0.08	0.09	0.13	0.05	0.04	0.01	0.08
sum	100.27	98.30	99.74	98.89	99.10	99.59	99.91	98.26	96.23	98.99	99.54	99.58	99.94
MgO/FeO*	0.35	0.35	0.35	0.34	0.35	0.33	0.29	0.32	0.34	0.36	0.35	0.34	0.34
CaO/Al_2O_3	0.81	0.81	0.79	0.79	0.79	0.81	0.86	0.82	0.84	0.79	0.79	0.80	0.82
FeO*/MnO	34.0	33.9	33.1	35.1	35.0	33.8	33.6	35.8	37.2	33.5	33.8	33.0	33.5

Table 3. Trace element dat for Antarctic eucrites in this study. All data in ppm except P_2O_5, TiO_2, and MnO as wt%. 'BHVO-2 PV' are the preferred values of this reference material used for calibration of the eucrite analyses.

	GRA 98098	EET 90029	BTN 00300	RKPA 80224	QUE 99658	QUE 97014	MAC 02522	MAC 041169	QUE 94484	PCA 91078	MET 01081	EE T90020	PCA 82502	BHVO-2 PV
P	418	239	244	156	276	321	348	409	571	223	161	59	330	1178
P_2O_5 wt%	0.096	0.055	0.056	0.036	0.063	0.073	0.080	0.094	0.131	0.051	0.037	0.013	0.076	0.270
Ti	3445	4544	3850	4294	4329	4472	5175	5368	6491	3128	3268	3741	4642	16886
TiO_2 wt%	0.57	0.76	0.64	0.72	0.72	0.75	0.86	0.90	1.08	0.52	0.55	0.62	0.77	2.82
Cr	1398	2017	2098	2025	2100	1959	1934	1910	2179	2386	2181	1985	2071	292
Mn	4282	4183	4291	4121	4176	4188	4724	4073	3777	4593	4262	4340	4358	1290
MnO wt%	0.55	0.54	0.55	0.53	0.54	0.54	0.61	0.53	0.49	0.59	0.55	0.56	0.56	0.17
Co	4.8	7.1	5.5	5.7	4.8	13.0	7.1	17.5	7.2	6.1	4.7	4.8	6.0	45.0
Ni	10.0	10.2	8.7	6.1	10.8	984.7	14.1	46.8	11.0	5.5	5.6	89.4	17.0	120
Ga	2.1	1.8	1.8	1.7	1.9	1.9	1.9	1.8	1.9	1.5	1.6	1.4	2.0	21.0
Rb	0.35	0.31	0.27	0.35	0.59	1.26	0.20	0.31	0.68	0.30	0.83	0.66	0.23	9.10
Sr	82.2	81.1	77.2	80.6	79.5	81.2	78.8	85.1	92.9	71.3	78.3	79.2	81.2	396
Y	21.9	19.5	16.1	17.4	17.9	19.1	19.9	24.0	30.8	11.2	10.9	12.4	12.9	24.9
Zr	77.6	54.8	41.6	51.8	51.4	53.7	56.6	76.5	104.9	30.7	24.7	37.9	54.7	175
Nb	4.43	4.11	3.15	3.86	3.91	4.08	4.14	5.49	8.28	2.40	2.09	2.71	4.11	18.8
Ba	43.2	35.8	32.7	32.4	35.2	36.6	35.8	43.4	62.7	28.7	21.8	33.3	37.4	131
La	4.31	3.58	3.01	3.16	3.32	3.58	3.33	4.65	6.45	1.99	1.74	1.08	2.01	15.2
Ce	10.94	8.65	7.29	8.38	7.95	8.77	8.42	11.49	16.08	4.36	3.65	2.38	9.74	37.5
Pr	1.59	1.31	1.08	1.16	1.20	1.29	1.22	1.72	2.42	0.65	0.55	0.35	0.82	5.29
Nd	8.03	6.63	5.44	5.89	6.03	6.47	6.18	8.69	12.10	3.38	2.88	2.04	4.06	24.5
Sm	2.48	2.07	1.74	1.87	1.93	2.08	2.02	2.74	3.78	1.13	1.00	0.87	1.39	6.07
Eu	0.67	0.67	0.60	0.65	0.66	0.66	0.63	0.73	0.85	0.54	0.60	0.61	0.65	2.07
Gd	3.44	2.96	2.44	2.66	2.77	2.89	2.94	3.81	5.13	1.61	1.48	1.51	1.89	6.24
Tb	0.62	0.53	0.44	0.47	0.49	0.52	0.54	0.68	0.89	0.30	0.28	0.30	0.37	0.96
Dy	3.89	3.43	2.82	3.11	3.18	3.34	3.53	4.32	5.61	1.98	1.88	2.09	2.42	5.31
Ho	0.83	0.73	0.61	0.66	0.68	0.71	0.75	0.91	1.18	0.43	0.41	0.48	0.52	0.97
Er	2.47	2.18	1.79	2.00	2.03	2.12	2.26	2.68	3.41	1.34	1.27	1.48	1.58	2.54
Yb	2.30	2.09	1.71	1.98	1.98	1.99	2.21	2.50	3.16	1.38	1.28	1.59	1.70	2.00
Lu	0.33	0.31	0.25	0.29	0.29	0.29	0.33	0.36	0.45	0.21	0.19	0.23	0.24	0.27
Hf	1.93	1.42	1.08	1.33	1.33	1.39	1.48	1.96	2.62	0.80	0.67	0.98	1.39	4.22
Ta	0.23	0.21	0.16	0.19	0.20	0.21	0.22	0.29	0.43	0.11	0.10	0.13	0.20	1.20
Th	0.56	0.42	0.33	0.40	0.38	0.39	0.42	0.56	0.84	0.22	0.15	0.06	0.43	1.26
U	0.16	0.09	0.08	0.05	0.10	0.10	0.13	0.15	0.22	0.07	0.04	0.02	0.10	0.42

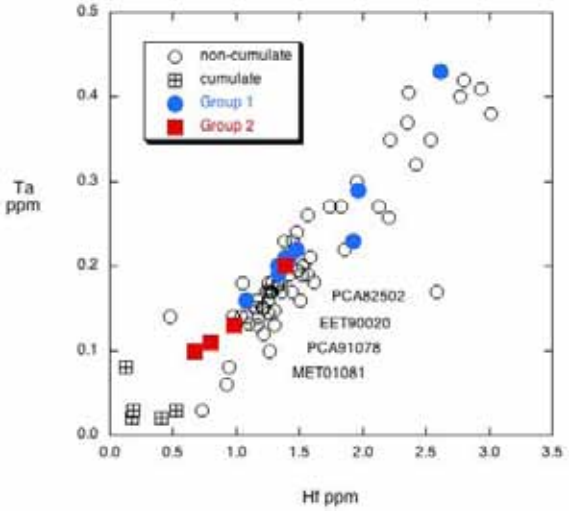


Figure 4. Incompatible trace element compositions of Antarctic eucrites (this study, red and blue filled circles) compared to literature data for noncumulate and cumulate eucrites [10, 11, 13].

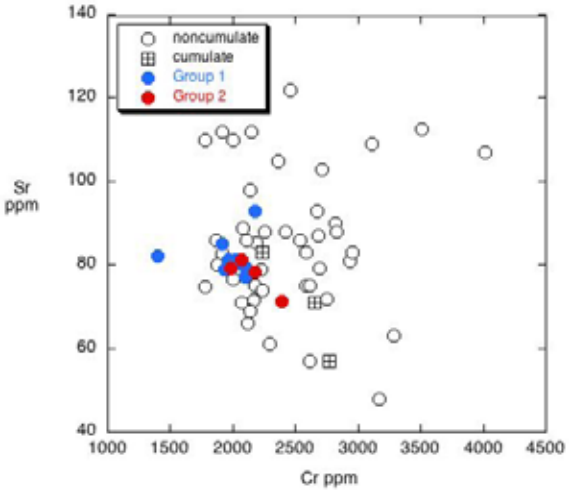


Figure 5. Trace elements compatible in pyroxene (Cr) and plagioclase (Sr) show limited compositional variations and no systematic differences between noncumulate and cumulate eucrites [10, 11, 13].

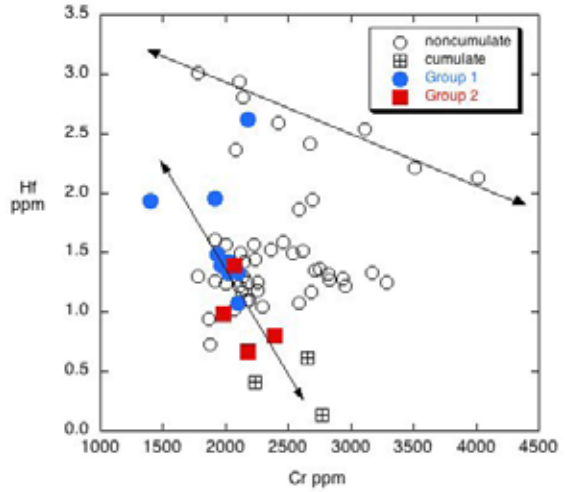


Figure 6a. Abundance of Cr. vs. Hf. Data from this study suggest negative correlations that link Group 1, Group 2, and cumulate eucrites whereas literature data [10, 11, 13]are more scattered. Arrows are suggestions of two trends.

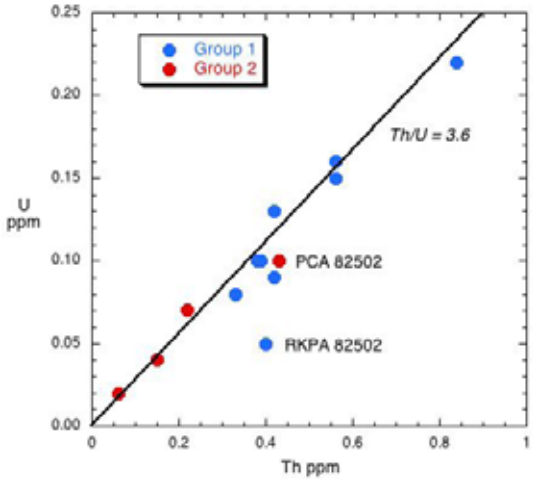


Figure 6b. Th vs. U abundances in Group 1 and Group 2 eucrites (this study). Most samples have Th/U ~3.6. A few samples have anomalously low U contents (e.g. RKPA 82502, Th/U = 8.0) probably indicating mobility of U in the terrestrial environment.

Igneous petrogenesis of eucrites

There are two main classes of petrogenetic models for explaining the compositions of basaltic eucrites: (1) variable degrees of partial melting (5-20%) of an undifferentiated parent body followed by fractional crystallisation of the parental magmas [12]; and (2) progressive tapping of residual liquids produced by crystallisation of a global magma ocean [11]. Distinguishing between these models is difficult because melting and crystallization are thermodynamically equivalent at equilibrium, and the small pressure gradient within the eucrite parent body (~500 km diameter if 4Vesta) produces similar mineral assemblages during both mantle melting and near-surface crystallization. Together these factors produce identical trends among the incompatible trace elements often used for modeling petrogenetic processes such as melting and crystallization.

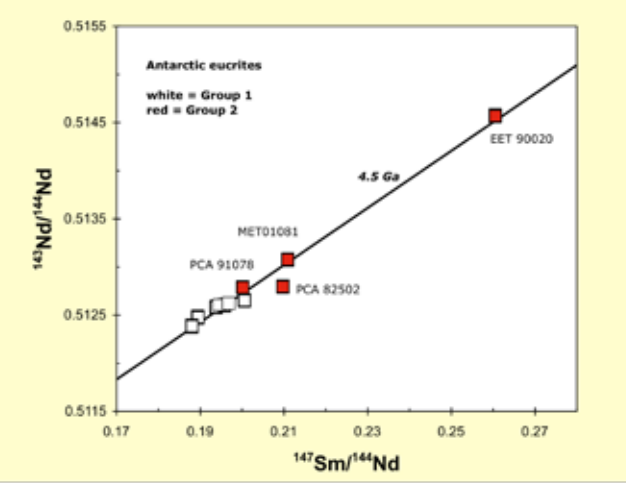


Figure 7. ^{147}Sm-^{143}Nd vs. $^{143}Nd/^{144}Nd$ isochron diagram for eucrites in this study. Group 1 eucrites have a small range of isotopic compositions reflecting their flat LREE patterns whereas Group 2 eucrites (red boxes) have a much broader range reflecting their LREE-depleted compositions.

	CHUR1 (Ga)	CHUR2 (Ga)
Group 1		
EET90029	5.20	6.27
QUE99658	4.23	5.36
QUE97014	2.99	3.25
MAC02522	0.58	0.77
RKPA80224	6.65	19.16
GRA98098	3.36	3.48
QUE94484	4.41	4.58
Group 2		
EET90020	4.56	4.53
PCA91078	6.33	5.71
PCA82502	1.82	1.83
MET01081	4.63	4.52
CHUR1		
143Nd/144Nd		0.512638
147Sm/144Nd		0.1966
CHUR2 Bouvier et al. (2008)		
143Nd/144Nd		0.512630

Table 4. Sm-Nd isotopic data of the eucrites presented as CHUR model ages calculated relative to the reference values given in the Table. CHUR values from Ref. [18].

The continuous compositional trends linking cumulates which have lost a melt fraction, with the noncumulate eucrites (Figs. 2, 4, 5) appear to support a primary role for fractional crystallization in generating most of the compositional variation within the noncumulate eucrites. On the other hand, the limited range of MgO/FeO (Fig. 2), the similar abundances of compatible elements in cumulate and noncumulate eucrites (Fig. 5), and the lack of strong correlations between compatible and incompatible trace elements (Fig. 6a) are difficult to explain by fractional crystallization and may be more consistent with a partial melting relationship [12]. Stolper [12] also notes that the MgO/FeO of pyroxenes in the cumulate eucrites imply equilibration with liquids that are more evolved (lower MgO/FeO) than observed in the noncumulate eucrites rather than representing early crystallizing phases from more primitive magmas.

Group 2 eucrites provide additional insight into igneous processes in the eucrite parent body. Major element compositions of Group 1 and Group 2 eucrites are indistinguishable (Table 2) and typical of noncumulate eucrites considered to be melts (Fig. 2). In contrast, the trace element characteristics of Group 2 samples, in particular their positive Eu anomalies (Figs. 3, 9), elevated Sr/Nd ratios (not shown), and low concentrations of incompatible trace elements (Figs. 4, 6), are more similar to cumulate eucrites. While Antarctic weathering may have modified the REE pattern of PCA 82502 (see previous section), the compositions of other three Group 2 eucrites appear to be unaffected by alteration.

Alternatively, we suggest that the compositions of at least some Group 2 eucrites reflect processes occurring on the eucrite parent body, and that eucrites with compositions similar to our Group 2 may be more common than previously recognized. Mittlefehldt and Lindstrom [17] noted that Antarctic eucrites with low REE also tend to have positive Eu anomalies, but considered these samples likely altered by terrestrial weathering, despite their lack of Ce anomalies. This is illustrated in Fig. 9 which shows that most of the noncumulate eucrites with positive Eu anomalies

(i.e., Eu/Eu* > 1) have low concentrations of the immobile trace element Ta, similar to the compositions of our Group 2 eucrites. This implies low intrinsic concentrations of other incompatible trace elements such as the LREE. Hf, Ta, and Th in these samples, and the possibility that at least some of the positive Eu anomalies in these eucrites may reflect primary characteristics of the eucrites rather than terrestrial alteration.

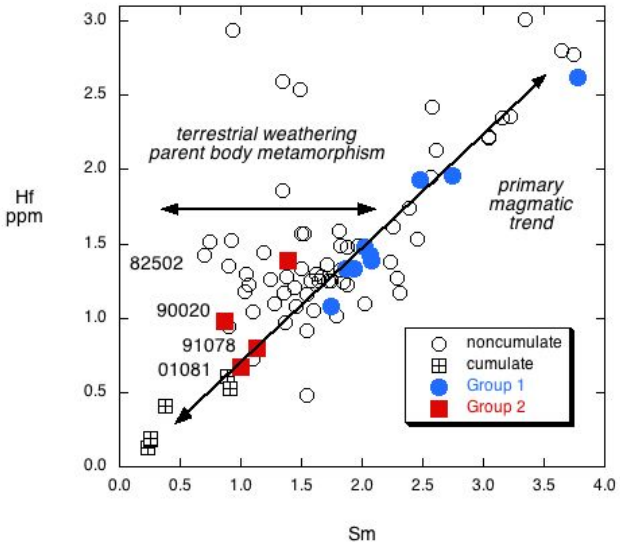


Figure 8. Samarium (Sm) vs. hafnium (Hf) concentrations in Group 1 (blue) and Group 2 (red) eucrites compared to literature data for cumulate and noncumulate eucrites. Sm and Hf should be strongly correlated due to their similar compatibilities during melting and crystallization. Group 1 eucrites have constant Sm/Hf ratios ~1.4 reflecting primary igneous compositions. Group 2 eucrites PCA 91078 and MET 01081 also have primary Sm/Hf ratios despite their low concentrations and LREE-depletions.

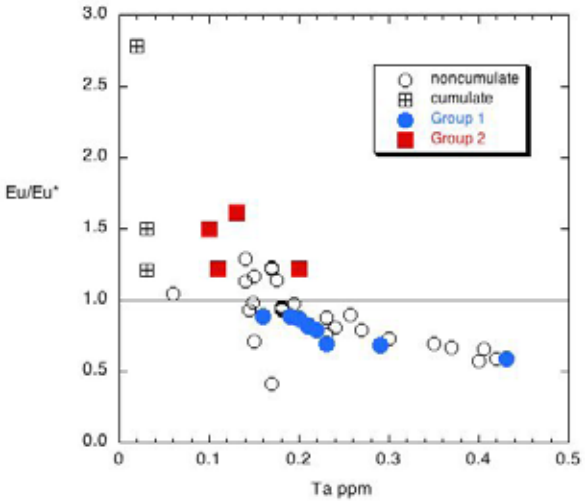


*Figure 9. Ta concentration vs. size of the Eu anomaly (=Eu_n/sqrt(Sm_n*Gd_n)) where subscript indicates chondrite-normalised concentration) in Group 1 and Group 2 eucrites of this study compared to literature data for noncumulate and cumulate eucrites [10, 11, 13].*

Neither partial melting of a chondritic mantle, nor fractional crystallization of eucrite basalt liquidus phases (pyroxene, plagioclase, spinel, ± olivine) [12] can produce LREE-depleted residual melts with positive Eu anomalies. The positive Eu anomalies (Figs. 3, 9) and elevated Sr/Nd ratios (not shown) of Group 2 eucrites clearly indicate a plagioclase signature, but magmatic accumulation into a primitive Group 1 type eucrite such as BTN 00300 or RKPA 80224 can be ruled out by the excessive amount of plagioclase addition that would be required to produce the Group 2 compositions (i.e. 20-70% of plagioclase with compositions like those in EET 90020) [7].

Alternative mechanisms for producing the major and trace element characteristics of Group 2 eucrites include (1) melting of a cumulate source enriched in plagioclase, and (2) extraction of small melt fractions from noncumulate eucrite magmas emplaced at moderate depths as dikes or sills. Yamaguchi et al. [7, 8] proposed local reheating and *in situ* melting of basaltic eucrites driven impacts into the hot early crust of Vesta to explain the textural variability and LREE-depleted patterns of some noncumulate eucrites. Melting of a plagioclase-enriched cumulate source would also transmit key trace element characteristics to the melt while major elements would remain buffered by phase proportions [12]. All Group 2 eucrites except for PCA 82502 are medium- to coarse-grained, indicating they crystallized at a moderate depth, possibly providing an opportunity for migration of a small residual melt fraction that would carry most of the highly incompatible elements such as the LREE, Hf, Ta, and Th while having only a minor effect on major elements and compatible trace elements such as Cr and Sr (Fig. 5). The mechanism of melt migration from a partially crystallised mush implies a possible genetic relationship between the Group 2 and cumulate eucrites, as suggested by some of the transitional compositional characteristics of the Group 2 eucrites described here (Figs. 4, 6), and may suggest an alternate interpretation of some cumulate eucrites such as Moore County. Whatever the mechanism, the Nd isotopic data show that the trace element fractionation occurred early in the history of the eucrite parent body. Additional study of the compositional and mineralogical characteristics of eucrites will be necessary to more fully understand their petrogenesis.

Acknowledgments

Reviews of the manuscript by Andrew Tomkins (Monash University) and Robert Pidgeon (Curtin University) are greatly appreciated. Partial funding for this work was received from the Australian Research Council (DP451432) and the Australian National University. We thank the NASA Johnson Space Center for the meteorite allocations.

References

1. Binzel, R. P. and Xu, S., "Chips off asteroid 4 Vesta: evidence for the parent body of basaltic achondrite meteorites", *Science*, Vol. 260, 1993, pp. 186–191.

2. Rayman, M.D., Fraschetti, T.C., Raymond, C.A. and Russell C.T., "Dawn: A mission in development for exploration of main belt asteroids Vesta and Ceres", *Acta Astronautica,* Vol. 58, 2006, pp. 605-616.

3. Reddy, V., Gaffey, M.J., Kelley, M.S. Nathues, A., Li, J.-Y. and Yarbrough R., "Compositional heterogeneity of Asteroid 4 Vesta's southern hemisphere: Implications for the Dawn mission", *Icarus*, Vol. 210, pp. 693-706.

4. Duke, M. and Silver, L., "Petrology of eucrites, howardites, and mesosiderites", *Geochim. Cosmochim. Acta*, Vol. 31, 1969, pp. 1637-1665.

5. *Basaltic Volcanism on the Terrestrial Planets (BVSP)*, Pergamon Press Inc, USA, 1981.

6. Mayne, R.G., McSween Jr., H.Y., McCoy, T.J. and Gale A., "Petrology of the unbrecciated eucrites", *Geochim. Cosmochim. Acta*, Vol. 73, 2009, pp. 794-819.

7. Yamaguchi, A., Taylor, G. J., Keil, K., Floss, C., Crozaz, G., Nyquist, L. E., Bogard, D. D.,

Garrison, D. H., Reese, Y. D., Weismann, H. and Shih, C.-Y., "Post-crystallization reheating and partial melting of eucrite EET 90020 by impact into the hot crust of asteroid 4Vesta 4.50 Ga", *Geochim. Cosmochim. Acta*, Vol. 65, 2001, pp. 3577–3599.

8. Yamaguchi, A., Barrat, J.A., Greenwood, R.C., Shirai, N., Okamoto, C., Setoyanagi, T., Ebihara, M., Franchi, I.A. and Bohn, M., "Crustal partial melting on Vesta: Evidence from highly metamorphosed eucrites", *Geochim. Cosmochim. Acta.*, Vol. 73, 2009, 7162-7182.

9. Bermingham, K., "Eucrites: an investigation into early planetary differentiation". Unpublished Honours Thesis. Dept. of Earth and Marine Sciences, Australian National University, 2007.

10. Kitts, K. and Lodders, K., "Survey and evaluation of eucrite bulk compositions", *Meteorit. Planet. Sci.*, Vol. 33, (1998), pp.197–213.

11. Barrat, J.A., Blichert-toft, J., Gillet, P. and Keller, F., "The differentiation of eucrites: the role of *in situ* crystallization", *Meteor. Planet. Sci.*, Vol. 35, pp. 1087-1100.

12. Stolper, E., "Experimental petrology of eucritic meteorites", *Geochim. Cosmochim. Acta*, Vol. 41, 1977, pp. 587–611.

13. Mittlefehldt, D.W. and Lindstrom, M.M., "Generation of abnormal trace element abundances in Antarctic eucrites by weathering", *Geochim. Cosmochim. Acta*, Vol. 55, 1991, pp. 77–87.

14. Jull, A.J.T., Cloudt, S. and Cielaszyk, E., "^{14}C terrestrial ages of meteorites from Victoria Land, Antarctica and the infall rates of meteorites" In *Meteorites: Flux with Time and Impact effects*, Grady, M.M., Hutchison. R., McCall G.J.H. and Rothery, D.A. (eds.), *Special Publications 140*, 1998, pp.75-91. Geological Society, London.

15. Freundel, M., Schultz, L. and Reedy, R.C., "Terrestrial ^{81}Kr-Kr ages of Antarctic meteorites", *Geochimica Cosmochimica Acta,* Vol. 50, 1986, pp. 2663-2613

16. Miura, Y., Nagao, K. and Fujitani**,** "^{81}Kr terrestrial ages and grouping of Yamato eucrites based on noble gas and chemical compositions", Geochim. Cosmochim. Acta, Vol. 57, 1993, pp. 1857-1866

17. Mittlefehldt, D.W. and Lindstrom, M.M., "Geochemistry of basaltic eucrites, and Hf and Ta as petrogenetic indicators for altered Antarctic eucrites", *Geochim. Cosmochim. Acta*, Vol. 67, 2003, pp. 1911–1935.

18. Bouvier, A., Vervoort, J.D., and Patchett J.P., "The Lu–Hf and Sm–Nd isotopic composition of CHUR: Constraints from unequilibrated chondrites and implications for the bulk composition of terrestrial planets", Earth Planet. Sci. Lett., Vol. 273, 2008, pp. 48–57.

Initial Results from a GIS-based Unsupervised Classification Study of the Martian Surface

Eriita Jones[1,2,3], Franklin Mills[3,4],Bruce Doran[3], Graziella Caprarelli[5] and Jonathan Clarke[2]

[1] *Research School of Astronomy and Astrophysics, Planetary Sciences Institute, Mount Stromlo Observatory, Australian National University, Cotter Road, Weston ACT, 2611.*
[2] *Mars Society Australia Inc.*
[3] *The Fenner School of Environment and Society, Building 48, Australian National University, Canberra ACT, 0200.*
[4] *Research School of Physics and Engineering, Mills Road, Australian National University, Canberra ACT, 0200.*
[5] *Department of Environmental Sciences, University of Technology, Sydney, PO Box 123, Broadway, NSW 2007, Australia*

Summary: Maps of thermal inertia-albedo units and thermal inertia-elevation units on Mars' surface have been generated by choosing thresholds that fit the strongest peaks in the histograms of these datasets. The units thus defined were then interpreted as distinct mixtures of materials on the surface, such as: bright fines, rock + bedrock and ice. We have conducted an initial classification of Thermal Emission Spectrometer (TES) night-time thermal inertia and TES albedo using a hard classifier. The methods used here are largely unsupervised and differ from those of previous studies. The aim of our study is to investigate what information can be obtained by utilising unsupervised classification algorithms to investigate the distribution of thermal materials on the surface of Mars. We find that unsupervised classification reveals additional structure in the clustering and spatial distribution of surface materials with moderate-low albedo and moderate-high thermal inertia. We highlight a number of regions such as Acidalia and Valles Marineris for future detailed studies of this type.

Keywords: Thermal inertia, albedo, TES, unsupervised algorithm, clusters, ISODATA, maximum likelihood.

Introduction

Mapping of remotely measured physical parameters on planetary surfaces provides insights into the nature of the surface materials and near-surface geology. Thermal inertia and albedo are particularly useful, because these physical parameters can be acquired easily by remote observation from repeated fly overs and for long periods of time. Putzig et al. (2005) extended the work done by Mellon et al. (2000) to derive seven two-dimensional groupings of pixels, which they termed 'thermo-physical units,' from global thermal inertia and albedo datasets generated from Mars Global Surveyor (MGS) Thermal Emission Spectrometer (TES) data (Christensen et al., 2001a). These authors mapped the spatial distribution of the thermo-physical units, which they interpreted as mixtures in various proportions of three principal end-members: dust, bedrock and ice. Putzig et al.'s (2005) map was biased towards the most distinctive thermal inertia and albedo information as their method involved choosing thresholds to encompass distinctive peaks in the histograms of the data. The aim of our study is to identify whether different or additional surface features of Mars could be identified through a

less deterministic method by applying an unsupervised classification algorithm to the data.

Thermal inertia is an indicator of the ability of a material to conduct and store heat and hence it parameterises the diurnal temperature profile of surface materials (Mellon & Jakosky, 1993). Ideally, thermal inertia is derived from observations of thermal emission at multiple times of day. However, this is often not possible for spacecraft data and leads to reduced spatial resolution. Thermal inertia can also be derived from single-time observations of surface thermal emission by comparing the observations with a subsurface heat conduction model that has a surface boundary condition given by:

$$(1) \qquad \frac{S}{R^2}(1-A)\cos(i) + F_{IR} + L\frac{\partial m}{\partial t} + I\sqrt{\pi/P}\,\frac{\partial T}{\partial Z'}\Big|_{Z'=0} = \varepsilon\sigma T_S^4$$

where S = solar flux, R = orbital radius of Mars, A = albedo of surface, i = the solar incidence angle, F_{IR} = thermal radiation from the atmosphere received at the surface, L = latent heat of sublimation of CO_2 (572893.8 J/kg); m = mass of CO_2 frost, I = thermal inertia, P = diurnal period, T = temperature, Z′ = depth below the surface normalized to the thermal skin depth, ε = emissivity of the soil surface, σ = the Stefan Boltzmann constant (5.670400×10^{-8} $Wm^{-2}K^{-4}$), T_S = surface temperature. The thermal skin depth is the depth at which temperature fluctuations within a material drop to 1/e of their surface value. The left-hand side of the equation is the summation of the input heat received at the surface and how this heat is distributed by the surface material. The right-hand side is the heat emitted by the surface. If all parameters in the heat conduction model are well constrained then thermal inertia can be estimated with good accuracy from the heat conduction model (Mellon et al., 2000; Mellon et al., 2004). Previous studies comparing thermal inertia obtained with this method to that from multiple time of day observations have found good agreement (Christensen and Malin, 1988).
For a pure material, thermal inertia is given by:

$$(2) \qquad I = \sqrt{k\rho C}$$

with thermal inertia units (tiu) of $K/m^2/K/s^{1/2}$ where k = thermal conductivity, ρ = density, C = volumetric heat capacity. Thermal inertia is most sensitive to variations in thermal conductivity, which on Mars is strongly related to both grain size and the degree of cementation of materials (Mellon at al., 2000). Density and heat capacity of materials on Mars vary by a factor of 2-3, however thermal conductivity varies by over 3 orders of magnitude (Neugebauer et al., 1971; Table 1). Fine-grained, loosely packed materials have lower conductivity and hence lower thermal inertia while larger particles, such as rocks and ices, have higher values of conductivity and thermal inertia. Materials with low thermal inertia respond quickly to temperature changes and will closely match the phase of diurnal temperature variations (Jakosky & Mellon, 2005a-b). High thermal inertia materials are slower to respond to temperature changes so their temperature profile lags behind the diurnal variations. Furthermore, the higher the thermal inertia of the material, the smaller the amplitude of its diurnal temperature variations, as the material is effectively dispersing the incident heat throughout its thermal bulk. At the spatial scale of spacecraft observations, the surface of Mars is typically a mixture of a range of materials. This complicates interpretation of apparent thermal inertia because the diurnal and seasonal apparent thermal inertia are very sensitive to the proportions of low and high thermal inertial materials within each pixel (Putzig & Mellon, 2007a,b). Furthermore, the apparent thermal inertia for a mixture differs fundamentally from that

of either pure component due to each component having a different temperature at any given time.

Albedo is the fraction of incident visible/near-IR solar radiation (0.3 – 2.9 μm) that is reflected by the surface (Christensen et al., 2001a). Bright materials have higher values of albedo and darker materials lower. For example, dark terrestrial soils have albedo ~ 0.05 compared to ~ 0.75 for fresh snow (Ahrens, 2006). As can be observed from Eqn. 1 materials with higher albedo values have a lower maximum temperature than dark materials (Mellon et al., 2000). The average albedo for Mars is ~ 0.20, ranging from 0.07 to 0.57, where an albedo value of 0.57 is similar to terrestrial bright sand, ice or old snow (Ahrens, 2006). Albedo varies strongly with the atmospheric redistribution of dust and seasonal condensation of CO_2 and H_2O (eg. Byrne et al. 2008). For example, the global dust storm in 2001 (Smith et al. 2002) brightened surface albedo values by as much as 0.1 (Putzig & Mellon, 2007) and likely induced large perturbations in day- and night-time surface temperatures (Wilson et al. 2007). Bright regions on Mars indicate fine-grained surface dust (Rogers & Bandfield, 2007; Bandfield & Smith, 2003). Dark regions correspond to mixtures of rocks, or duricrust with smaller proportions of dust and ices. Low albedo regions have been found to have strongly homogeneous mineral compositions (Rogers & Bandfield, 2007), consisting of a mixture of minimally weathered basalt (surface type 1; Bandfield & Hamilton, 2000), altered basalt (surface type 2; Karunatillake & Squyres, 2006) and hematite (Christensen et al., 2001b).

Albedo and thermal inertia are only partially independent as both are related to thermal conductivity. As shown in Eqn. 2, thermal inertia is related to thermal conductivity by a power law. Thermal conductivity is also related to particle size with larger particles typically sharing a larger surface area with their neighbours and hence having a higher bulk thermal conductivity (for grains < 1mm under conditions on Mars; Jakosky, 1986). Albedo is correlated with particle size, with surfaces composed of larger particles typically being darker (Shkuratov *et al.* 1999) due to specular reflection. Hence albedo and thermal conductivity of surface materials are generally negatively correlated. Furthermore the albedo and thermal conductivity of a given soil varies inversely with the soil moisture, with increased water within a soil lowering the albedo but raising the conductivity (Wang *et al.,* 2005). Hence on Mars, high thermal inertia materials (eg. rocks) predominantly span a lower range of albedo values than small grained, low thermal inertia materials (eg. dust, sand). A scatterplot of global thermal inertia and albedo values on Mars (see Fig. 4 of Putzig et al., 2005) reveals the complex relationship between these variables, showing a weak negative correlation between albedo and thermal inertia for values of the latter < ~ 250 tiu (grain size ~ 170 μm) and a weakly positive correlation for thermal inertia values > ~ 250 tiu.

Thermal inertia and albedo provide two partially independent constraints on the nature of surface materials, and their combined use for mapping surface features on Mars was first demonstrated by Palluconi & Kieffer (1981). Putzig et al. (2005) chose thresholds for their thermo-physical classes that isolated the strongest modes in the histograms of the two datasets (Fig. 1) The histogram of Mars' global albedo contains three strong peaks – at 0.15, 0.23 and 0.27 (binwidth of 0.01). Thermal inertia shows two strong peaks – at 55 tiu and at 225 tiu (binwidth of 5). The former is caused by dust on Mars; the latter includes contributions from a range of materials. The methodology used in Putzig et al. (2005) in defining the unit boundaries based on the histogram peaks enables a clear detection of regions that are predominantly fine-grained surface dust because they have a very high albedo and a very low thermal inertia. However the

mixtures of materials that contribute to the two peaks at lower albedo are degenerate in thermal inertia, ie., although the materials can be distinguished through their albedo they each cover a similar range in thermal inertia (eg. Units B, C, E and G in Putzig et al. 2005, Table 1). Hence it is not clear if the somewhat arbitrary thresholds chosen in the study reflect the true underlying data structure in albedo-thermal inertia space. It is for this reason that we have employed an unsupervised algorithm that iteratively groups pixels based on minimising their Euclidean separation in albedo-thermal inertia space.

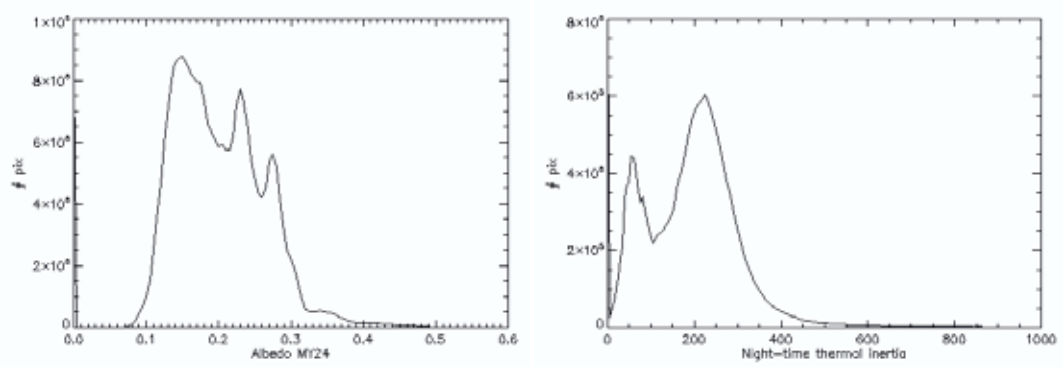


Figure 1: Histograms of the global albedo (left) and night-time thermal inertia (right) data sets. Horizontal axis has been cropped to focus on peaks. Both datasets are multi-modal and contain 2.592 x 10^7 pixels.

Unsupervised classification techniques have recently been applied to Martian Digital Terrain Models to identify topographic features such as craters and valley ridges (eg. Stepinski *et al.,* 2005, 2009a,b) however they have not been utilized to identify the distribution of thermally distinct surface materials. Mapping the thermal behavior of surface material on Mars has many interesting applications, one of which is constraining subsurface temperature profiles which can then be utilized to estimate the range of depths at which liquid water or brine could exist (eg. Jones & Lineweaver, 2010).

Methods

The technical specifications of the Mars Global Surveyor Thermal Emission Spectrometer used to collect the albedo dataset and the data from which thermal inertia is derived are described in detail in Christensen et al. (2001a). The two global datasets used in this study – Mars Year 24 (MY24) albedo and night-time thermal inertia - were obtained from http://lasp.colorado.edu/inertia/2007/albedo.html (2007 Global MGS-TES Albedo Maps) and http://lasp.colorado.edu/inertia/2007/ (2007 Global MGS-TES Thermal Inertia Maps). The MY24 data were selected because MY24 had minimal localised dust storm events (Cantor et al., 2002) and generally a lower dust optical depth (the atmosphere was more transparent) than MY25 and 26 (Tamppari et al., 2008; Smith 2004), which are also available. This means that the derived albedo values for MY24 will be more representative of the mean surface materials and less affected by scattering due to atmospheric dust. The datasets have dimensions of 7200 x 3600 pixels which correspond to a resolution of 0.05˚ per pixel or ~3 km^2. The data sets have been bilinearly interpolated between the MGS orbital tracks with real data constituting ~93% of the thermal inertia map and ~35 % of the albedo map (Putzig pers. comm.). The uncertainty in albedo values is approximately ± 0.01 (Christensen et al., 2001a) and the uncertainty associated with each derived night-time thermal inertia value is estimated

to be < 10% (Putzig & Mellon, 2005), lower than for daytime measurements (error < 17%). The TES sensing depth is on the order of several centimeters (Mellon et al., 2000), so TES provides information only on surficial levels of the crust. Data was analysed using: off-the-shelf commercial software, ArcMap 9.2, for description, mapping and visualization of the distribution of the classes; and IDL for data manipulation prior to importing into ArcMap and analysis. Classification algorithms were used to find 7 distinct clusters in the data, so that a direct comparison with the 7 clusters defined by Putzig et al. (2005) could be undertaken.

Walker *et al.* (1986) gave one of the earliest presentations of the technique of unsupervised classification by using a scatter plot of pixel values in two spectral bands to separate land cover features. Unsupervised classification is now broadly used in the interpretation of terrestrial remote sensing data (Hall et al., 1995). The technique involves using a clustering algorithm to group pixels in parameter space that have similar values within each measurement parameter (parameters are typically red and near-infrared reflectances in a terrestrial LANDSAT image but are albedo and thermal inertia in this study) without *a priori* knowledge of the surface materials present. This is often necessary with remotely sensed data of a large or complex area as the actual or optimum number of natural clusters in the data is not known and the amount of data is so vast that it must be reduced before field data can be collected to define the link between remotely sensed pixel values and surface land cover. By grouping observations and minimizing the differences between members in a group, the spatial locations that are most similar can be identified (Duda & Hart, 1973; Murray & Estivill-Castro, 2001). Once a cluster dataset has been produced, each cluster can then be interpreted as a mixture of materials on the surface with the values for each cluster used to determine some of the attributes of those surface materials (Jupp *et al.,* 1986).

There are two fundamental caveats to the technique of unsupervised classification. Firstly, the number of clusters to be found by the classification algorithm must be chosen by the user (e.g. Milligan & Cooper, 1985) and hence may not reflect the full complexity of natural patterns within the data. Secondly, there may be several acceptable clusterings of the dataset, depending upon the purposes of the application. Despite this, cluster outputs produced by unsupervised classification using the maximum likelihood decision rule applied in this study have been applied successfully in terrestrial remote sensing (e.g. Belward et al., 1990). Studies using LANDSAT-TM, MSS and ETM imagery which use field data to assess the accuracy of the unsupervised classification, report accuracies (i.e. the percentage of pixels within a cluster that have been identified as belonging to that cluster) of 60 – 90% (Miller & Yool, 2002; Sader *et al.,* 1995; Murthy *et al.,* 2003). Whilst comparable field data is not readily available for analysis of remotely sensed Mars data, the generic unsupervised classification approach and associated techniques used in terrestrial applications are transferrable.

The albedo and thermal inertia datasets were both normalized before classification by scaling the max and min to the range [0,1]. The algorithms used to generate the 7 two-dimensional clusters were ISODATA (Iterative Self-Organizing Data Analysis Technique) and Maximum Likelihood Classification (MAXLIKE) run through ArcGIS 9.2, using the calling commands ISOCLUSTER and MLCLASSIFY, respectively. ISODATA is an unsupervised training algorithm which uncovers the statistical patterns inherent in the data. It is useful when little is known about the data prior to classification. ISODATA works by examining a subset of the data (every 5th pixel in this study) to provide candidate clusters which are then fed into the clustering algorithm. These candidate

clusters are analogous to the training data that would be used in a supervised classification (for example, if the signatures of a certain type of feature in the dataset were known before classification). The user inputs the maximum number of clusters to generate (7 in this study) and then the algorithm iteratively passes through the dataset defining clusters until either the maximum number of iterations is reached (specified by user; Ball & Hall, 1965; Richards, 1986) or there is little change in clusters between iterations (Swain 1973). Pixels are determined to be similar if they have similar values in each of the n-input parameters (Tou & Gonzalez, 1974). Calculating the parameter distance using the Euclidean distance equation (Richards, 1986; p.190[1]), the ISODATA algorithm assigns pixels to a cluster if the parameter distance between the pixel and the cluster centre is less than the distance to all other cluster centers. To do this, an initial vector with a mean value for N clusters (where N is the maximum number of final clusters specified by the user) is defined so that the N initial arbitrary cluster centers are uniformly spaced along the multidimensional diagonal ensuring that the initial assignment of cluster centers is not biased to the extrema of the input data. The location of the initial cluster means is not important so long as enough iterations are allowed (500 in this study) for the clusters to become stable. On the first pass through the data, each pixel is compared with the vector and assigned to the cluster whose centre is closest in parameter distance. After the first pass through the data the cluster centers are redefined and new centers are calculated by taking the sample mean of pixels assigned to each cluster in the previous step. The process then continues iteratively. The final number of clusters can be less than N when: (a) a minimum cluster size S is specified (in this study S = 30, corresponding to 0.0001 % of the dataset) so that any clusters consisting of fewer cells will be eliminated at the end of an iteration; or, (b) clusters close in parameter space and with similar statistical values become merged during the iterative process The assignment of clusters is independent of the contiguity of the pixels in the spatial frame.

The candidate clusters provided by ISODATA are described in a signature file that provides the mean and covariance matrix for each parameter in each cluster and is used to train the classification algorithm MAXLIKE. MAXLIKE assumes that the distribution of each cluster will be multivariate normal (depending on the number of parameters in the classification, hence bivariate normal[2] in this study) so that each cluster can be characterized by the statistical parameters provided in the signature file for the corresponding training cluster. Each pixel in the dataset is then assigned to a single cluster according to the Boolean decision rule based on probabilities from the normal distributions. A pixel X belongs to the cluster S_i if the probability of X belonging to S_i is greater than the probability of X belonging to S_j, for all j≠ i'. Hence each pixel is assigned to the cluster to which it has the highest probability of being a member, irrespective ofthe actual probability of membership. It is assumed that all clusters have an equal probability of occurring. The MAXLIKE algorithm works best with a normal distribution of data (Kloer, 1994). If the histograms of the input parameters are strongly non-normal (as in this study; Fig. 1) the algorithm can overestimate χ^2 (chi-squared) classification errors (Benson & Fleishman, 1994). This problem is minimized with an increasing

[1] The parameter distance is calculated by the Euclidean distance equation: $D_j = \left(\sum_{i=1}^{dim}(x_i^2 - s_i^2)\right)^{1/2}$ where dim = the number of data dimensions (in this study, dim = 2), x = data vector, s = cluster centre vector.

[2] The bivariate normal distribution has probability density function: $P(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} exp\left[-\frac{z}{2(1-\rho^2)}\right]$ where $z = \frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}$, $\rho = \frac{v_{xy}}{\sigma_x\sigma_y}$ and μ = mean, σ = stanard deviation, v = covariance (Wackerly *et al.,* 2002; p. 268).

number of data points in the total sample and has been found to be negligible if the number of data points is > 400 (Harlow et al., 1985).

The validity of the clusters defined by MAXLIKE can be quantified through the χ^2 distribution with degrees of freedom equal to the number of input parameters (Swain & Davis, 1978; Richards, 1986). This provides a measure of the classification confidence by giving the percentage chance for each pixel that it has been assigned to the correct cluster, based on the separation distance between the pixel value and the cluster mean in the multiparameter space. Essentially, if the number of pixels in a given cluster is plotted against their parameter distance from the cluster mean this will ideally follow a χ^2 distribution (Richards, 1986 p. 174). Thus, if a threshold confidence value of 97.5% is imposed, the least likely 2.5% of pixels in the cluster (those with χ^2 values greater than 7.4 for 2 degrees of freedom) will be rejected (Wackerly et al. 2002 , p. 794). The χ^2 statistics are generally applied to independent variables. Although the covariance of the (normalized) albedo and thermal inertia datasets used in this study is non-zero (cov = 0.002), it is small compared to their individual variances (0.015 and 0.011, respectively). The MAXLIKE algorithm run through ArcMap outputs a confidence raster based on the χ^2 distribution (see Results), which provides the level of confidence with which each cell was classified. For example, cells with a value of 5 have a 95% chance of having been correctly classified. Furthermore, the presence of clear unfragmented spatial patterns in the classification map provides anecdotal evidence that the classifier is producing a reasonably accurate model (Gahegan & West, 1998).

Interpreting thermal inertia & albedo

The atmosphere of Mars is never dust free and the surface, globally, is dominated by particulate matter (Bandfield *et al.,* 2000). Understanding the thermal inertia and albedo signatures of these fines is key to interpreting the thermo-physical units. Albedo is strongly correlated with the degree of fine, bright dust coverage (Kieffer, 1973; Ruff & Christensen, 2002). Surfaces on Mars covered by fine, bright dust have an albedo > 0.25 and a low value of thermal inertia because of the low thermal conductivity of small particles. A combination of infrared spectral analysis and studies of analogue materials at Martian surface pressures showed that dust on Mars typically has diameters less than ~ 40 micrometers (Christensen, 1986) which correspond to a thermal inertia < 65 tiu (Jakosky 1986). Larger, coarse, unbonded particulates from 100 - 10^4 micrometres dominate the low albedo regions of Mars (Christensen & Moore, 1992). For grains up to ~ 1000 micrometres across (thermal inertia ~ 200) laboratory studies show that thermal conductivity increases linearly with particle size (Jakosky, 1986) and, hence, the thermal inertia generally increases with the square root of grain size. Larger particulates have a thermal inertia ~ 400 tiu (Jakosky, 1986). Thus, the combination of thermal inertia and albedo can be used to identify surfaces that are completely covered by fine-grained dust. Identifying larger darker fines (mm – cm) is more difficult, however, as both (i) their albedo is consistent with rocks and duricrust and (ii) their thermal inertia is consistent with a sub-pixel-scale mixture fine dust with high thermal inertia materials (such as rocks). Hence, intermediate values of thermal inertia do not have a unique interpretation given they can result from a range of mixtures of materials. Materials such as rock and ice are more easily identified through thermal inertia, as values above 2500 tiu can only be regions that are pure or almost pure rock or ice. Representative values drawn from laboratory, in-situ, observational, and modelling analyses for the thermal inertia of some materials on Mars are shown in Table 1.

Table 1: Representative properties of some materials on Mars (taken from Jones & Lineweaver, 2011 in prep.)

Material	Grain size (mm)	Thermal conductivity ($Wm^{-1}K^{-1}$)	Density (kgm^{-3})	Heat capacity ($Jkg^{-1}K^{-1}$)	Thermal diffusivity (m^2s^{-1})	Calculated TI (Eqn. 2)	References
Dust	<< 1 x 10^{-3} -0.1	0.001	1000	800	1.3 x 10^{-9}	28	Mellon & Phillips 2001 Jakosky 1986 Karunatillake et al. 2010
Sand	0.1 - 10	0.1	1750	800	7.1 x 10^{-8}	374	Mellon & Phillips 2001 Heldman *et al.* 2005 Murphy *et al.* 2009 Jakosky 1986
Duricrust	Highly variable; sand size particles cemented to form larger grains	0.3 (up to 2)	1750	800	2 x 10^{-7}	648	Murphy *et al.* 2009 Piqueux & Christensen 2009 Fergason et al. 2006
Icy soil	N/A	2.5	2018	1040	1.2 x 10^{-6}	2294	Mellon *et al.* 2004 Sizemore & Mellon 2006
Rock	> 40	2.5 (range generally 1.5-4.5)	2900	800	1 x 10^{-6}	2408	Clauser & Huenges, 1995 Turcotte *et al*. 2002 Sizemore & Mellon 2006 Golombek et al. 2005 Mellon & Phillips 2001
H_2O ice	N/A	3.4 (at -110 °C ; 2 at 0°C)	928	1310	1.3 x 10^{-6}	2044	Clauser & Huenges 1995 Titus *et al.* 2003

Each TES pixel is of order 3 km^2, which makes it unlikely any pixel contains only one type of surface material, so each of our thermo-physical units is interpreted as a 'mixture' of materials on the surface of Mars. Our interpretation of the material components of each cluster is based on matching the albedo and thermal inertia peaks in Fig. 2 to the estimated parameters of materials on Mars shown in Table 1. The main components of the mixtures of materials in each of the 7 clusters are given in Table 2.

As expected, the main source of uncertainty in interpreting the results of the calculations is the degeneracy of the thermal inertia-albedo data (see Introduction). For example, the peak in thermal inertia at ~ 250 tiu (Clusters 1-4 in Fig. 2) is consistent with either pure coarse- grained unconsolidated particles, such as sand, or a mixture of duricrust (~600 tiu), dust (~50 tiu) and sand (~ 300 tiu). An albedo below ~ 0.2 implies the absence of significant amounts of fine, bright dust and hence the thermal inertia is interpreted as a weighted average of the amount of duricrust and sand where the weights vary.

Results

The seven thermo-physical units defined in this study are mapped in Fig. 3 and details are given in Table 2. A pattern of global 'enveloping' can be seen in the spatial occurrence of the 7 clusters, mirrored on either side of the Equator (Fig. 3). This corresponds to a decrease in albedo with increasing distance from the Equator through the sequence of clusters 5 ⟶ 4 ⟶ 3 ⟶ 2 ⟶ 1. The mean thermal inertia also increases through this sequence - with the exception of Cluster 3 – and the values indicate that the fraction of bright dust and fine grained sand is decreasing with progression through the cluster sequence. However, Cluster 3 is an exception to this as it has a thermal inertia peak corresponding to fine, unconsolidated material, associated with a decreased albedo relative to Cluster 4. This may be a detection of dark fines. Cluster 3 is also present near the South Pole where it corresponds to Putzig et al.'s Class D (light blue) which was interpreted by these authors as 'dark dust'. If this interpretation is correct, it covers an extensive area of the surface of Mars, predominantly in the southern polar terrain but also infilling Hellas Basin, Promethei Terra and the albedo feature west of Elysium.

The map in Fig. 2 has associated error maps provided in Fig. 3. These maps indicate which regions were most likely to have been classified accurately by the unsupervised algorithm. 58% of pixels in Fig. 2 are most likely classified correctly, as they fall within the 50% confidence interval (Fig. 3). The 50% confidence interval contains pixels from each of the seven thermo-physical units indicating that that the general surface locations in which these clusters occur is fairly well constrained. The core spatial patterns of each cluster are maintained in both the 50% and 75% confidence interval maps. The most accurate pixels, classified with $>$ 95% confidence, comprise 15% of the map. These pixels are located primarily equatorwards of $\pm$ 70° and very few of them belong to cluster 6.

Table 2: Some details of the 7 thermo-physical units from this study in the form: mean, standard deviation.

Cluster	Thermal inertia (tiu)	Albedo (%)	Interpretation
1	260, 70	13, 1	Duricrust, rocks, sand
2	230, 70	16, 1	Duricrust, sand
3	180, 80	20, 1	Duricrust, sand, dark fines
4	200, 80	24, 1	Duricrust, sand, dust
5	80, 50	28, 2	Dust
6	410, 340	35, 6	Ice
7	3670, 920	30, 7	Ice, rocks

Both the thermo-physical map of Putzig *et al.* (2005) and the map produced in this study closely match previous determinations of the global spatial pattern of Mars' fine, bright dust cover (Ruff & Christensen, 2002, Bandfield *et al.,* 2000). There are broad similarities between our map and that of Putzig *et al.* (2005), particularly in the equatorial regions of Mars where small unconsolidated particles dominate. The spatial boundary of Cluster 5 in this study closely matches Putzig *et al.* (2005) Class A (blue) and is dominated by dust (Ruff & Christensen, 2002). The distribution of Cluster 4 is very similar to Putzig *et al.* (2005) Class C (green), although Cluster 4 appears to be always associated with Cluster 3, which was not differentiated in the Putzig *et al.* (2005) study (it was part of Class B, yellow). Other differences between the results of this work and Putzig *et al.* (2005) are that our Classes 1 and 2 are not differentiated in Putzig *et al.* (2005), as their classification incorporated them into their broad mid-latitude Class B. Our Cluster 6 occurs at the poles and incorporates high-latitude data published in Putzig and Mellon (2007; hence not included in the Putzig *et al.* (2005) map). Possible anomalously high thermal inertia values occurring at the North Pole are contained in Putzig *et al.* (2005) Class E and our Cluster 7.

In synthesis, both maps classify bright regions (eg. Tharsis, Arabia and Elysium) as being dominated by the same thermo-physical unit, which is interpreted as fine, bright dust deposits at least several centimetres thick.. The main differences occur in low albedo regions (eg. Acidalia and Sinus Meridiani), which will be investigated in follow-up studies.

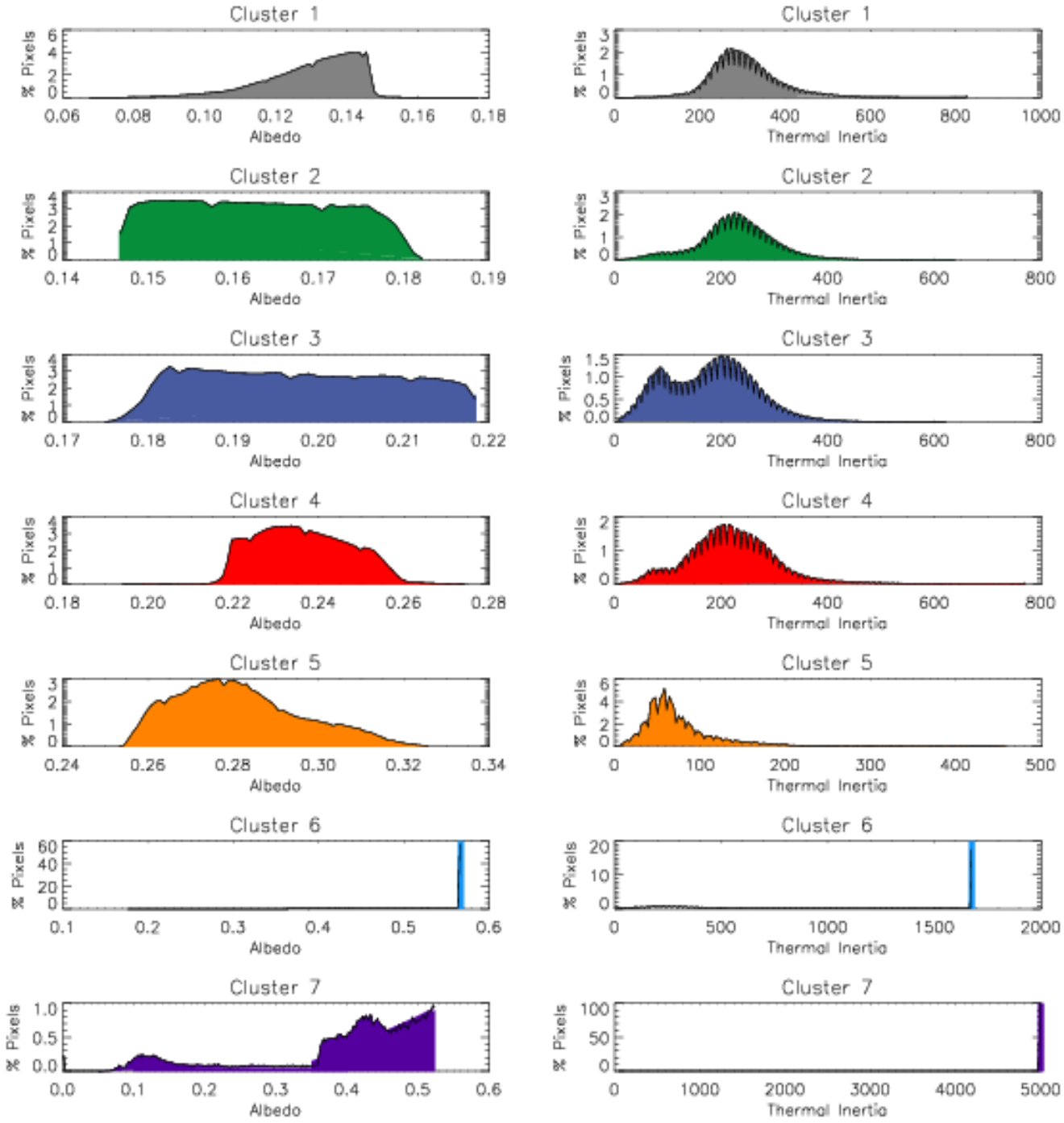


Figure 2: Histograms of albedo and thermal inertia for the 7 thermo-physical units.

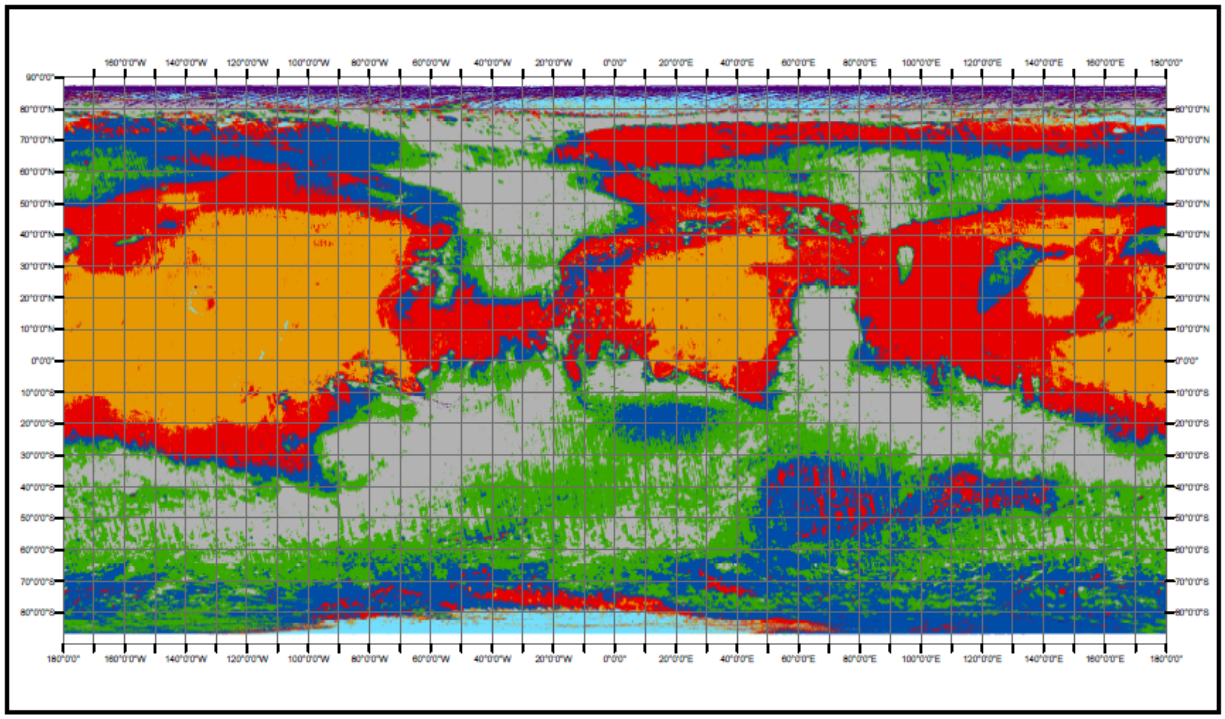


Figure 3: Unsupervised classification of TES albedo and night-time thermal inertia into seven thermo-physical units. Details of the units are given in Table 2 and histograms in Figure 2.

Conclusions

We have used albedo and night-time thermal inertia data of the surface of Mars and an unsupervised classification method to identify seven mappable thermo-physical units. Our results refine previous studies (Putzig et al., 2005; Putzig and Mellon 2007a,b) of the distribution of rock, dust and ice on the surface. We interpreted the thermo-physical units by comparing the peaks in the albedo and thermal inertia histogram of each unit to representative values of materials on Mars. Our results agree with previous studies in the equatorial region where bright fines dominate. However, we find evidence of additional structure in the distribution of surface materials at higher latitudes, particularly in areas of low albedo and moderate-high thermal inertia. Further studies will need to be undertaken to resolve the reasons for the differences and determine the nature of the surface material at high latitudes. This work significantly demonstrates that unsupervised classification can detect potentially important structure in surface materials on Mars at a higher spatial resolution than that provided in previous studies.

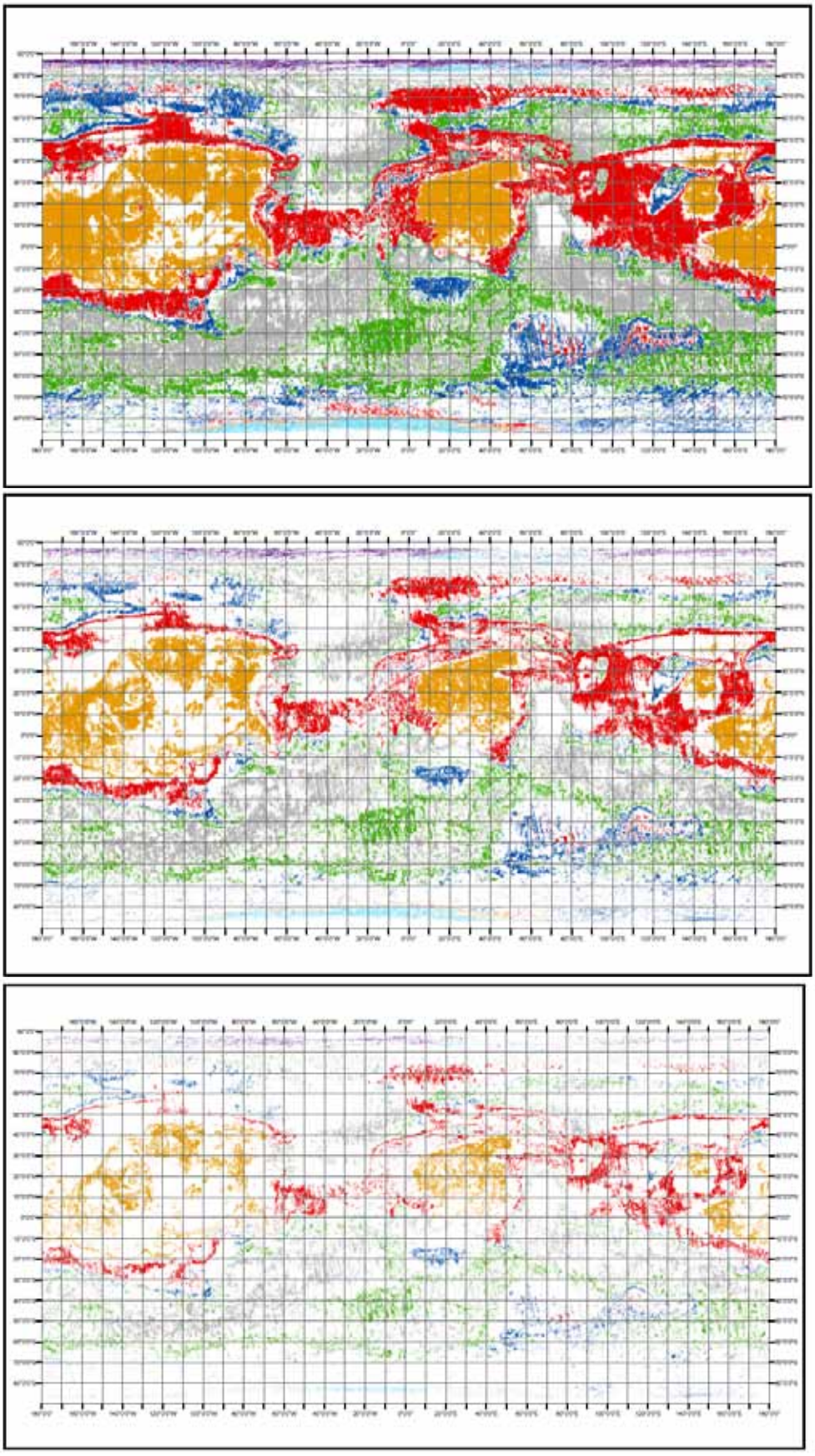


Figure 4: Classification error map at 50%, 75% and 95% confidence intervals. Pixels are either coloured to match their cluster assignment in Figure 3, or coloured white. White pixels have either a: > 50% chance of being classified into the wrong cluster (top); > 25% chance of being misclassified (middle); or > 5% chance of being misclassified (bottom; see details on χ^2 distribution in text).

Acknowledgements: ArcGIS 9.3 Online Help, Topics: "How Maximum Likelihood Classification Works" and "How Iso Cluster Works", ESRI, 2010.
ERDAS Inc., "Unsupervised Training" and "Evaluating Classification" published in ERDAS Field Guide (3rd Ed), Atlanta, 1994, pp. 240-247 and pp. 270-278 respectively.

References

1. Ahrens, C. "Meteorology Today: An Introduction To Weather, Climate, and The Environment (8th Ed.)", Cengage Learning, USA, 2007, p. 44.
2. Ball, G.H., Hall, D.J., "ISODATA, A Novel Method of Data Analysis and Pattern Classification", Stanford Research Institute, California, 1965.
3. Bandfield, J. L., Hamilton, V. E., Christensen, P.R., "A Global View of Martian Surface Compositions From MGS-TES", *Science*, Vol. 287, 2000, pp. 1626-1630.
4. Bandfield, J. and Smith, M., "Multiple Emission Angle Surface-Atmosphere Separations of Thermal Emission Spectrometer Data", *Icarus*, Vol. 161, 2003, pp. 47-65.
5. Belward, A.S., Taylor, J.C., Stuttard, M.J., et al., "An Unsupervised Approach To The Classification of Semi-Natural Vegetation From Landsat Thematic Mapper Data. A Pilot Study On Islay", *International Journal of Remote Sensing*, Vol. 11, 1990, pp. 429-445.
6. Benson, J., Fleishman, J.A., "The Robustness of Maximum Likelihood and Distribution-Free Estimators To Non-Normality In Confirmatory Factor Analysis", *Quality and Quantity,* Vol. 28, 1994, pp. 117-136.
7. Byrne, S., Zuber, M.T., Neumann, G.A., "Interannual and Seasonal Behaviour of Martian Residual Ice-Cap Albedo", *Planetary and Space Science*, Vol. 56, 2008, pp. 194-211.
8. Cantor, B.A., Malin, M., Edgett, K.S., "Multiyear Mars Orbiter Camera (MOC) Observations of Repeated Martian Weather Phenomena During The Northern Summer Season", *Journal of Geophysical Research*, Vol. 107, 2002, E001588.
9. Christensen, P.R., "Regional Dust Properties On Mars: Physical Properties, Age and History", *Journal of Geophysical Research*, Vol. 91, 1986, pp. 3533-3545.
10. Christensen, P.R., Malin, M.C., "High Resolution Thermal Imaging of Mars", *Lunar and Planetary Sciences Conference,* Vol. 19, 1988, pp. 180-181.
11. Christensen, P.R., Moore, H.J., "The Martian Surface Layer", in *Mars*, Kasting, J., Eds, 1992, pp. 686-729.
12. Christensen, P.R., Bandfield, J.L., Hamilton, V.E., and 23 others, "Mars Global Surveyor Thermal Emission Spectrometer Experiment : Investigation Description and Surface Science Results", *Journal of Geophysical Research*, Vol. 106, 2001a, pp. 23823-23871.
13. Christensen, P.R., Morris, R.V., Lane, M.D., et al., "Global Mapping of Martian Hematite Mineral Deposits: Remnants of Water-Driven Processes On Early Mars", *Journal of Geophysical Research*, Vol. 106, 2001b, pp. 23873-23885.
14. Clauser, R., Huenges, E., "Thermal Conductivity of Rocks and Minerals" in Rock Physics and Phase Relations – A Handbook and Phase Relations, Ahrens, T., Eds, American Geophysical Union, 1995, pp. 105-126..
15. Duda, R.O., Hart, P.E., "Pattern Classification and Scene Analysis", *Wiley Publishers*, New Jersey, 1973.
16. Fergason, R.L., Christensen, P.R., Bell III, J.F., et al., "Physical Properties of the Mars Exploration Rover Landing Sites As Inferred From Mini-TES-Derived Thermal Inertia", *Journal of Geophysical Research,* Vol. 111, 2006, E02S21.
17. Gahegan, M., West, G., "The Classification of Complex Geographic Datasets: An Operational Comparison of Artificial Neural Networks", *Geocomputation Conference*, 1998. Accessible at: http://www.geocomputation.org/1998/61/gc_61.htm
18. Golombek, M.P., Arvidson, R.E., Bell, J.F., and 13 others, "Assessment of Mars Exploration Rover Landing Site Predictions", *Nature,* Vol. 436, 2005, pp. 44-48.

19. Hall, F.G., Townshend, J.R., Engman, E.T., "Status of Remote Sensing Algorithms For Estimation of Land Surface State Parameters", *Remote Sensing of Environment*, Vol. 51, 1995, pp. 138-156.
20. Harlow, L., Chou, C.P., Bentler, P., "Performance of Chi-Square Statistic With ML, ADF, and Elliptical Estimators for Covariance Structures", *Annual Meeting of the Psychometric Society*, Toronto, Canada, 1986.
21. Heldman, J., Toon, O., Pollard, W., et al., "Formation of Martian Gullies By The Action of Liquid Water Flowing Under Current Martian Environmental Conditions", *Journal of Geophysical Research*, Vol. 110, 2005, E05004.
22. Jakosky, B. M., "On the Thermal Properties of Martian Fines", *Icarus*, Vol. 66, 1986, pp. 117-124.
23. Jakosky, B.M., Mellon, M.T., Varnes, S., et al., "Low-Latitude Neutron Distribution: Possible Remnant Near-Surface Water Ice and A Mechanism For Its Recent Emplacement", *Icarus*, Vol. 175, 2005a, pp. 58-67.
24. Jakosky, B.M., Mellon, M.T. , Varnes, S., et al., "Erratum to Low-Latitude Neutron Distribution: Possible Remnant Near-Surface Water Ice and A Mechanism for Its Recent Emplacement", *Icarus*, Vol. 175, 2005b, pp. 58-67.
25. Jones, E.G., Lineweaver, C.H., "To What Extent Does Life 'Follow the Water'?", *Astrobiology*, Vol. 10, 2010, pp. 349-361.
26. Jones, E.G., Lineweaver, C.H., "Phase diagram of the Potential Martian Biosphere", 2011, in preparation.
27. Jupp, D.L., Walker, J., Penridge, L.K., "Interpretation of Vegetation Structure In Landsat MSS Imagery: A Case Study In Disturbed Semi-Arid Eucalypt Woodlands. Part 1. Field Data Analysis", *Journal of Environmental Management,* Vol. 23, 1986, pp. 19-33.
28. Karunatillake, S., Squyres, S.W., Taylor, G.J. and 12 others, "Composition of Northern Low-Albedo Regions of Mars: Insights From the Mars Odyssey Gamma Ray Spectrometer", *Journal of Geophysical Research (Planets)*, Vol. 111, 2006, E03S05.
29. Karunatillake, S., McLennan, S., Herkenhoff, K.E., "Regional and Grain Size Influences On the Geochemistry of Soil At Gusev crater, Mars", *Journal of Geophysical Research,* Vol. 115, 2010, E00F04.
30. Kieffer, H.H., Chase, S.C, Miner, E., et al., "Preliminary Report On Infrared Radiometric Measurements From The Mariner 9 Spacecraft", *Journal of Geophysical Research*, Vol. 78, 1973, pp. 4291-4312.
31. Kloer, B.R., "Hybrid Parametric/Non-Parametric Image Classification", *ASPRS/ACSM Annual Convention*, 1994, pp. 307-316.
32. Mellon, M.T., Jakosky, B.M., "Geographic Variations in the Thermal and Diffusive Stability of Ground Ice on Mars", *Journal of Geophysical Research*, Vol. 98, 1993, pp. 3345-3364.
33. Mellon, M.T., Jakosky, B.M., Kieffer, H.H., Christensen P.R., "High-Resolution Thermal Inertia Mapping from the Mars Global Surveyor Thermal Emission Spectrometer", *Icarus*, Vol. 148, 2000, pp. 437-455.
34. Mellon, M.T., Phillips, R.J., "Recent Gullies On Mars and The Source of Liquid Water", Journal of Geophysical Research, Vol. 106, 2001, pp. 1-15.
35. Mellon, M.T., Feldman, W.C., Prettyman, T.H., "The Presence and Stability of Ground Ice In The Southern Hemisphere of Mars", *Icarus*, Vol. 169, 2004, pp. 324-340.
36. Miller, J.D., Yool, S.R., "Mapping Forest Post-Fire Canopy Consumption In Several Overstory Types Using Multi-temporal Landsat TM and ETM data", *Remote Sensing of Environment*, Vol. 82, 2002, pp. 481-496.
37. Milligan, G.W., Cooper, M.C., "An Examination of Procedures For Determining The Number of Clusters In A Data Set.", *Psychometrika*, Vol. 50, 1985, pp. 159-179.
38. Murray, A.T., Estivill-Castro, V., "Cluster Discovery Techniques For Exploratory Spatial Data Analysis", *International Journal of Geographical Information Science*, Vol. 12, 1998, pp. 431-443.
39. Murphy, N., Jakosky, B.M., Mellon, M.T., Budd, D., "Thermophysical Properties of Martian Duricrust Analogs.", *40th Lunar and Planetary Sciences Conference,* Vol. 40, 2009, abstract #1420.
40. Murthy, C.S., Raju,, P.V., Badrinath, K.V., "Classification of Wheat Crop With Multi-temporal Images: Performance of Maximum Likelihood and Artificial Neural Networks", *International Journal of Remote Sensing*, Vol. 24, 2003, pp. 4871-4890.
41. Neugebauer, G., Munch, G., Kieffer, H., et al., "Mariner 1969 Infrared Radiometer Results: Temperatures and Thermal Properties of the Martian Surface", *Astronomical Journal,* Vol. 76, 1971, pp. 719-749.

42. Palluconi, F.D., Kieffer, H.H., "Thermal Inertia Mapping of Mars From 60˚S to 60˚N", *Icarus*, Vol. 45, 1981, pp. 415-426.
43. Piqueux, S., Christensen, P.R., "A Model of Thermal Conductivity For Planetary Soils: 2. Theory For Cemented Soils", *Journal of Geophysical Research*, Vol. 114, 2009, E09006.
44. Putzig, N.E, Mellon, M.T., Kretke, K.A., Arvidson, R.E., "Global Thermal Inertia and Surface Properties of Mars From The MGS Mapping Mission", *Icarus*, Vol. 173, 2005, pp325-341.
45. Putzig, N.E., "Thermal Inertia and Surface Heterogeneity On Mars", *PhD Dissertation*, 2006.
46. Putzig, N.E. & Mellon, M.T., "Thermal Behaviour of Horizontally Mixed Surfaces On Mars", *Icarus*, Vol. 191, 2007a, pp. 52-67.
47. Putzig, N.E, Mellon, M.T., "Apparent Thermal Inertia and The Surface Heterogeneity of Mars", *Icarus*, Vol. 191, 2007b, pp. 68-94.
48. Richards, J.A., "Remote Sensing Digital Image Analysis – An Introduction", *Springer-Verlag*, Berlin, 1986, pp. 175-204.
49. Rogers, D.A., Bandfield, J.L., Christensen, P.R., "Global Spectral Classification of Martian Low-Albedo Regions With Mars Global Surveyor Thermal Emission Spectrometer (MGS-TES) Data", *Journal of Geophysical Research (Planets)*, Vol. 112, 2007, E02004.
50. Ruff, S.W., Christensen,P.R., "Bright and Dark Regions of Mars: Particle Size and Mineralogical Characteristics Based On The Thermal Emission Spectrometer", *Journal of Geophysical Research*, Vol. 107, 2002, E001580.
51. Sader, S.S., Ahl, D., Liou, W.S., "Accuracy of Landsat-TM and GIS Rule-Based Methods For Forest Wetland Classification", *Remote Sensing of Environment*, Vol. 53, 1995, pp. 133-144.
52. Shkuratov, Y., Starukhina, L., Hoffman, H., Arnold, G., "A Model of Spectral Albedo of Particulate Surfaces: Implications For Optical Properties of The Moon", *Icarus*, Vol. 137, 1999, pp. 235-246.
53. Sizemore, H.G., Mellon, M.T., "Effects of Soil Heterogeneity On Martian Ground-Ice Stability and Orbital Estimates of Ice Table Depth", *Icarus*, Vol. 185, 2006, pp. 358-369.
54. Skinner, J.A., Jr., Hare, T.M., Tanaka, K.L., "Digital Renovation of the Atlas of Mars 1:15,000,000-Scale Global Geologic Series Maps", Lunar and Planetary Sciences Conference, Vol. 38, 2006, abstract #2331.
55. Smith, M.D., Conrath, B.J., Pearl, J.C., Christensen, P.R., "Thermal Emission Spectrometer Observations of Martian Planet-Encircling Dust Storm 2001A", *Icarus*, Vol. 157, 2002, pp. 259-263.
56. Smith, M.D., "Interannual Variability in TES Atmospheric Observations of Mars During 1999-2003", *Icarus*, Vol. 167, 2004, pp. 148-165.
57. Stepinski, T., Vilalta, R., "Digital Topography Models For Martian Surfaces", *IEEE Geoscience and Remote Sensing Letters*, Vol. 2, 2005, pp. 260-264.
58. Stepinski, T.F., Bagaria, C., "Segmentation-Based Unsupervised Terrain Classification For Generation of Physiographic Maps", *IEEE Geoscience and Remote Sensing Letters*, Vol. 6, 2009a, pp. 733-737.
59. Stepinski, T.F., Mendenhall, M.P., Bue, B.D., "Machine Cataloging of Impact Craters On Mars", *Icarus*, Vol. 203, 2009b, pp. 77-87.
60. Swain, P.H., Davis, S.M., "Remote Sensing: The Quantitative Approach", *McGraw-Hill International*, Sheffield, UK, 1978.
61. Swain, P.H., "Pattern Recognition: A Basis For Remote Sensing Data Analysis", LARS information note #111572, Purdue University, 1973.
62. Tamppari, L.K., Smith, M.D., Bass, D.S., Hale, A.S., "Water-Ice Clouds and Dust In The North Polar Region of Mars Using MGS TES Data", *Planetary and Space Science*, Vol. 56, 2008, pp. 227-245.
63. Tanaka, K.L, Skinner, J.A., Hare, T.M., "Geologic Map of the Northern Plains of Mars", US Geological Survey, Scientific Investigations Map 2888, 2005.
64. Titus, T.N., Kieffer, H.H., Christensen, P.R., "Exposed Water Ice Discovered Near The South Pole of Mars", *Science*, Vol. 299, 2003, pp. 1048-1051.
65. Tou, J.T., Gonzalez, R.C., 'Pattern Recognition Principes', *Adisson Wesley*, London, 1974.
66. Turcotte, D.L., Scherbakov, R., Malamund, B.D., Kucinskas, A.B., "Is the Martian Crust Also The Martian Elastic Lithosphere?", *Journal of Geophysical Research*, Vol. 107, 2002, E001594.
67. Wackerly, D.D., Mendenhall III, W., Scheaffer, R.L. "Mathematic Statistics With Applications (6th ed.)" *Thomson Learning*, Pacific Grove, 2002.

68. Walker, J., Jupp, D., Penridge, L., Tian, G., "Interpretation of Vegetation Structure In Landsat MSS Imagery: A Case Study In Disturbed Semi-Arid Eucalypt Woodlands. Part I: Field Data Analysis.", *Journal of Environmental Management*, Vol. 23, 1986, pp 19-33.
69. Wang, K., Wang, P., Liu, J., Sparrow, M., Haginoya, S., Zhou, X., 'Variation of Surface Albedo and Soil Thermal Parameters With Soil Moisture Content At A Semi-Desert Site On The Western Tibetan Plateau.', *Boundary-Layer Meteorology*, Vol. 116, 2005, pp. 117-129.
70. Wilson, R.J., Neumann, G.A., Smith, M.D., "Diurnal Variation and Radiative Influence of Martian Water Ice Clouds", *Geophysical Research Letters*, Vol. 34, 2007, L02710.

Using dual-polarised L-band SAR and optical satellite imagery for land cover classification in Southern Vietnam: comparison and combination

Hai Tung Chu[1], Linlin Ge[2] and Xin Wang[3]

[1,2,3]School of Surveying and Spatial Information Systems
The University of New South Wales
Sydney NSW 2052, AUSTRALIA

Keywords: optical, Synthetic Aperture Radar, like/cross polarisation, classification, accuracy, Artificial Neural Network, Support Vector Machine.

Summary: Synthetic Aperture Radar (SAR) and optical images contain different types of land cover information. Moreover, whereas like-polarisation SAR images are predominately influenced by surface backscattering properties, the volume backscattering characteristics of surfaces are well represented by SAR cross-polarisation data. Hence, combined use of like- and cross-polarisation data could enhance the land cover mapping process.

This paper investigates the use of multi-temporal dual-polarised L-band ALOS /PALSAR in combination with SPOT satellite images for land cover classification in Southern Vietnam. The classification processes were implemented for different datasets, including single-type image datasets and different combined datasets. Incorporation of textural information was also attempted. The non-parametric classifiers, namely the Artificial Neural Network (ANN) and Support Vector Machine (SVM) were applied. This study demonstrates the advantages of using multi-temporal dual-polarised SAR data and the significant improvement of classification accuracy using a combination of multi-temporal dual-polarised ALOS/PALSAR and SPOT images for land cover classification.

Introduction

Land cover mapping is one of the most important applications of remote sensing data. Both optical and Synthetic Aperture Radar (SAR) satellite images are employed for this purpose.
The optical and SAR satellite images are obtained using different regions of the electromagnetic spectrum, therefore they sense different characteristics of the ground objects over the same imaged area. The optical images such as Landsat TM, SPOT, and MODIS are recorded in the visible to near-infrared regions, providing information on reflectivity, transmissivity, and absorption capability of land cover features. On the other hand, the SAR data acquired from the microwave region of the spectrum can provide information concerning roughness, dielectric content (water content) and structures of the illuminated surfaces. Thus, a combination of SAR and optical images could provide complimentary information and lead to improved land cover classification results. Numerous studies have been undertaken using this combination approach, with different datasets and classification techniques (e.g. Kuplich et al. 2000, Chust et al. 2004,

Erasmi and Twelve 2009, Kim and Lee 2005, Huang et al. 2007, Sheoran et al. 2005). Although the reported results vary considerably, most authors claimed that integration of SAR and optical data improved classification performance. In Kim and Lee (2005), land cover features were classified using several combinations of Landsat ETM+ and Radarsat images. The overall classification accuracy using combined datasets was improved to 74.6% as compared to an accuracy of 69.35% using only Landsat ETM+ data. Erasmi and Twele (2009) evaluated the synergy of ENVISAT/ASAR data and Landsat ETM+ data for land cover classification in Central Sulawesi, Indonesia using the maximum likelihood classification algorithm. It was revealed that the integration of ASAR and Landsat TM images increased classification accuracy significantly, and that the combination of like-polarised time series SAR and optical images produced the best results. Kuplich et al. (2000) claimed that the combination of Landsat TM and SAR images increased the classification accuracy for pasture and planted forest classes. In Sheoran et al. (2009) combinations of a quad-polarisation PALSAR and Landsat TM images gave very high classification accuracy of over 90%, which is better than using either single dataset on its own.

The radar systems which send and receive the same kind of polarised signal (either HH or VV) are referred to like-polarisation systems, while radar systems which send one kind of polarised signal and receive another kind are referred to as cross-polarised systems (either HV or VH).

The like-polarised SAR systems are sensitive to the surface scattering mechanism, which occurs when the radar signal strikes and backscatters from surfaces such as rocks, bare ground, built-up areas, etc. In this type of scattering, the polarisation of the transmitted and reflected signals are unchanged. In other words, there are no, or very little, de-polarisation effects in the backscattered energies. By contrast, the cross-polarised SAR systems are more sensitive to the volume scattering mechanism, where radar signals are depolarised by surface materials and changes in polarisation of returned energies are received by antennas. The volume scattering often occurs when the radar signal strikes tree canopy, whereby the radar pulses interact with various tree components such as leaves, branches, and so, which cause depolarisation of the radar signal. Therefore, use of both like- and cross-polarised (or dual-polarised) data has potential for enhancing the separability of land cover classes.

Textured data is an important characteristic of satellite imagery, which gives unique information on spatial patterns, arrangement and variation of ground features. Incorporation of textured data with spectral information for land cover mapping has been implemented by many researchers and very encouraging results were reported (e.g. Huang et al. 2007, Herold et al. 2005). The most common textured measures are computed using the Grey Level Co-occurrence Matrix (GLCM) (Lu and Weng, 2007).

Beside input data, the classification techniques that are used are extremely important for land cover mapping. A broad range of classification algorithms has been developed and applied for classifying remote sensing data. The traditional parametric classifiers, such as the Maximum Likelihood (ML) classifier, is widely used (Huang et al. 2002, Waske and Braun, 2009) because they provide acceptable accuracy with relatively little computation effort. The major limitation of these algorithms is the assumption of normal distribution of input data – which is not often the case for remote sensing data (Waske and Benediktson, 2007). This makes it difficult for such parametric classifiers to integrate multi-source data, such as combining SAR and optical data. Unlike parametric classifiers, the non-parametric classifiers such as Artificial Neural Network (ANN) or Support Vector Machine (SVM), do not constrain their application to the assumption of normal distribution, and are therefore often more appropriate for classifying remote sensing data.

A number of studies have confirmed that non-parametric techniques produce better classification accuracy than parametric techniques (e.g. Dixon and Candade 2008, Huang et al. 2002, Kavzoglu and Mather 2003). In Huang et al. (2002), ML, ANN and SVM classifiers were tested for

classifying Landsat TM images. Results showed that the ANN and SVM techniques generally gave significantly higher accuracies than the ML algorithm. The SVM technique resulted in better accuracy than the ANN technique when using seven input variables, while ANN gave higher accuracies than SVM in the case of only three input variables. In Dixon and Candade (2008), both SVM and ANN classification techniques resulted in better accuracy than ML classification, while SVP and ANN classification had comparable quality of results. Nevertheless, few studies have applied ANN or SVM classification techniques for classifying combinations of different kinds of satellite imagery data, such as mixing SAR and optical images.

The major objective of this study was to investigate the potential of using multi-temporal, dual-polarised L-band ALOS /PALSAR images in combination with SPOT multi-spectral satellite images for land cover classification in Southern Vietnam, using the ANN and SVM classification techniques. The classification process was carried out for different single-type and combination datasets, in order to evaluate their performance for mapping land cover features.

Study area and data used

The study area was located along the Saigon River in Southern Vietnam, centred at the coordinate 106^{o} 46 E; 10^{o} 48' 30'' N. The area is part of the Ho Chi Minh City, and a small part of the Dong Nai province (Figure 1). The terrain is very flat with the main land cover features being water surface, vegetation, bare land and urban areas.

The SPOT 2 multi-spectral image obtained on January 10, 2008, and five ALOS/PALSAR dual-polarised (HH/HV) images acquired for the period from June 08, 2007 to June 10, 2008, were used for this study (Table 1).

Table 1: ALOS/PALSAR images for the study area

Satellite/Sensor	Track - Frame	Acquisition dates	Polarisation	Orbit	Spatial Resolution
ALOS/PALSAR	477_20	08 Jun 2007	HH/HV	Ascending	12.5 m
		08 Sep 2007	HH/HV	Ascending	12.5 m
		09 Dec 2007	HH/HV	Ascending	12.5 m
		March 2008	HH/HV	Ascending	12.5 m
		10 Jun 2008	HH/HV	Ascending	12.5 m

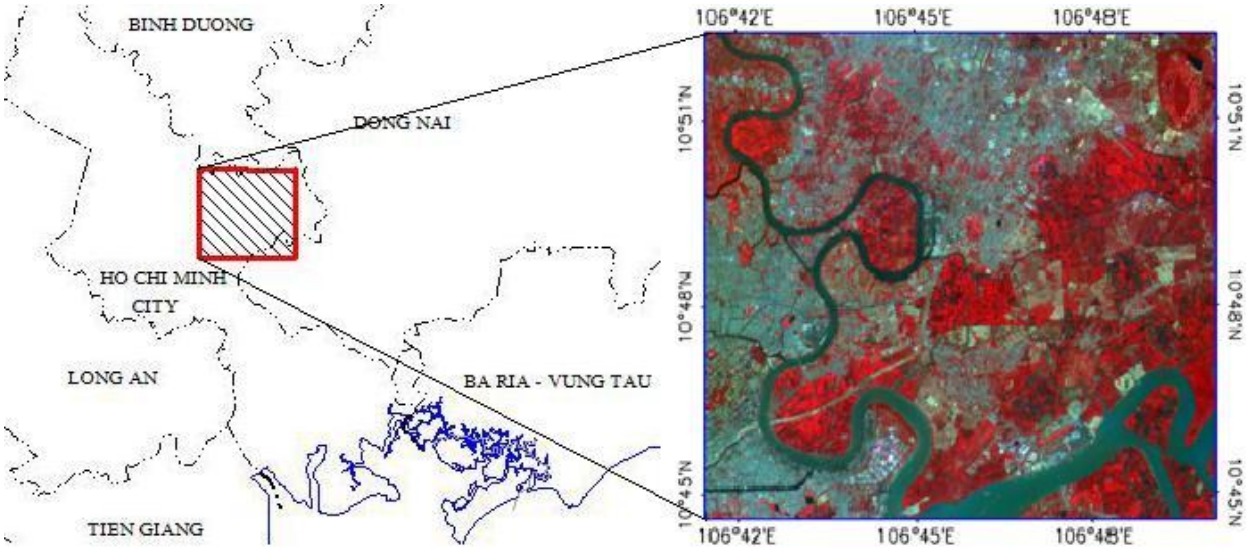


Fig 1. Location of the study area

Methodology

ALOS/PALSAR and SPOT 2 multi-spectral images were geo-rectified to map coordinates (WGS84, UTM projection, zone 48) with sub-pixel accuracy. Since the terrain is very flat, no DEM was required for ortho-rectification. All data were re-sampled to 12.5m of pixel size. The adaptive Enhanced Lee filter with window size of 3x3 was applied to remove speckle noise from the SAR images. PALSAR backscatter values were converted to decibel (db).

In this study, textured data was employed for multi-date SAR images only. The possible use of textural information for both SAR and optical images will be the subject of future work. Major objectives of using textural information was to evaluate possibilities to improve land cover mapping accuracy by combining multi-date PALSAR images with their textural information.
GLCM texture measures were extracted from the First Principle Component (PC1) of each five-date PALSAR like- (HH) and cross- (HV) polarisation datasets. Four GLCM texture measures were employed, namely Variance, Homogeneity, Entropy and Correlation. Since there is no preferred direction, average of texture measures generated at 8 different direction including 0^o, 45^o, 90^o, 135^o, 180^o, 225^o, 270^o, 315^o were computed and integrated with SAR backscatter data for classification. Various window sizes were tested for textured generation, including 3x3, 5x5, 7x7, 9x9, 11x11 and 13x13. At each window size the incorporation of GLCM texture measures with multi-date PALSAR images were carried out using single, two, three and all four measures to generate different combined datasets. These datasets will be used for classification with both SVM and ANN classifiers. Two sets of textural features which provide highest accuracy while integrated with five-date HH and HV images respectively, will be identified and combined with five-date dual (HH+HV) images for classification.
The Multi Layer Perception (MLP) model with a back propagation learning algorithm is a common type of ANN applied in remote sensing, and was used for this study. The networks consisted of three layers, including input, hidden and output layer – since the use of more than three layers does not improve accuracy significantly (Mills et al. 2004). Each input feature, such as SPOT image bands, were introduced to the network by one input layer node. Land cover classes were presented in the output layer nodes. The hidden layer nodes were automatically

structured and pruned by the ENVI software used for classification (Argany et al. 2006). The sigmoid function was used as an activated function:

$$F(x) = \frac{1}{1+e^{-x}} \quad (1)$$

$$x = wp + b \quad (2)$$

Where x represented total input at a node, p denotes input values, w is weight of input and b is bias.

The parameters that were used for the ANN classification processes were: Number of iterations: 1000; Learning rate: 0.1; Training momentum: 0.9; Training RMS Exit Criteria: 0.1.

The SVM algorithms transform the training data in the lower dimensional input space into a higher dimensional feature space using a kernel function. This process allows non-linear separated data in the input space to be linearly separable in the high dimensional feature space. The SVM using the Radial Basic Function kernel, equation (3) below, was selected to classify land cover features in the study site because of its robustness (Kavzoglu and Colkesen, 2009). The cross-validation technique was applied to determine the optimal parameters, including γ (width of a kernel) and C (penalties coefficient) for the SVM classifier.

Radical Basis Function (RBF) kernel is:

$$K(x,y) = \exp(-\gamma\|x-y\|^2) \quad (3)$$

Where x, y represent a training samples and their class labels, respectively in a feature space, γ is width of a kernel.

Six land cover classes were identified for classification. These classes were: Water Surface (WF), Bare Land (BL), Dense High Urban structures (DHU), Low Flat Urban structures (LFU), High Dense Vegetation (HDV), and Low Spared Vegetation (LSV). The spectral properties and backscatter signatures of these features in the SPOT 2 multi-spectral and multi-temporal PALSAR images are shown in Figure 2.

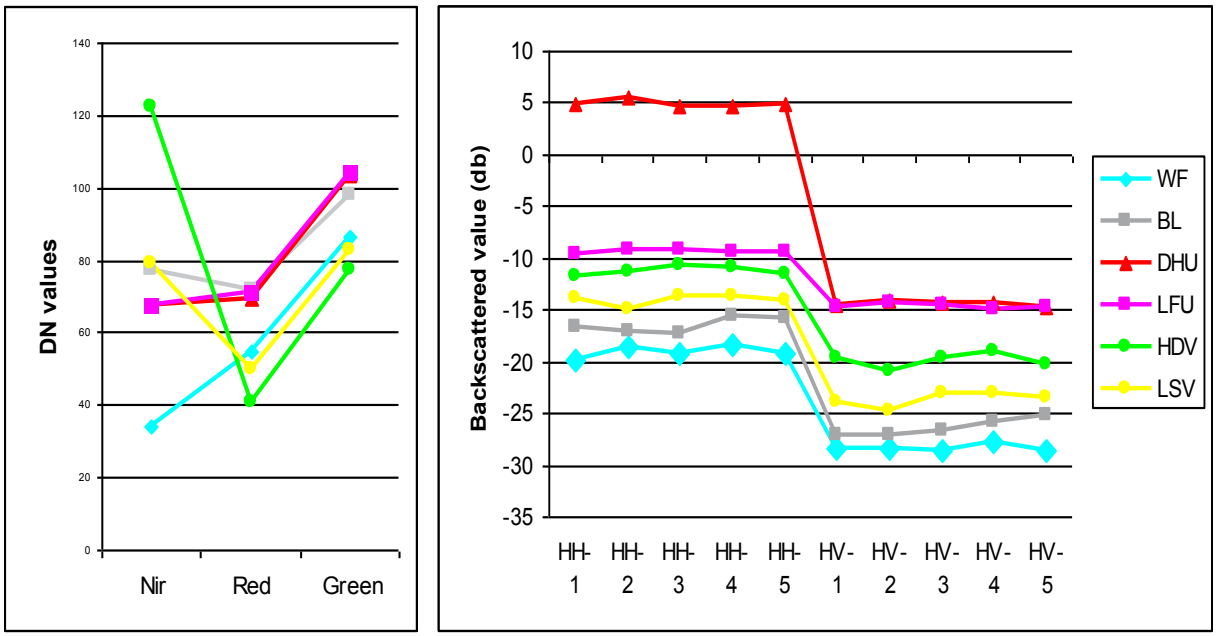


Fig. 2: Land cover feature characteristics in SPOT 2 multi-spectral and multi-temporal PALSAR images

The validation data was randomly selected based on manual interpretation with reference to old land use maps created in 2005 and field checking.

Results and discussion

The overall classification accuracy for the ANN and SVM classifiers over different datasets is summarised in the Table 2 and figure 3.

Table 2: Land cover classification accuracy of different datasets
(Note: the ANN classifier cannot classify a single image)

Datasets	Overall classification accuracy (%)	
	SVM	ANN
Single-date PALSAR HH polarised images (1Date_HH)	~49.35	-
Single-date PALSAR HV polarised image (1Date_HV)	~45.53	-
Single-date PALSAR dual-polarised images HH+HV (1Date_Dual)	~57.80	~54.99
Five date PALSAR HH polarised images (5Date_HH)	58.12	56.34
Five date PALSAR HV polarised images	54.00	50.89
Five date PALSAR dual HH+HV polarised images (5Date_Dual)	65.51	68.16
SPOT2 multi-spectral images (SPOT 2 XS)	69.43	70.08
SPOT2 multi-spectral + single PALSAR dual-polarised HH+HV images (XS +1Date_Dual)	83.60	86.55
SPOT2 multi-spectral + five-date PALSAR HH polarised images (XS + 5Date_HH)	89.27	88.67
SPOT2 + five-date PALSAR HV polarised images (XS + 5Date_HV)	75.31	72.92
SPOT2+ five-date PALSAR dual HH+HV polarised images (XS + 5Date_Dual)	88.31	89.16

Single-date, single-polarised SAR images provided very poor classification accuracy – approximately 49.35% for HH and 45.53% for HV polarised images using the SVM classifier. The classification of a single SAR image was not possible for the ANN classifier. The single-date, dual-polarised image resulted in an increase of approximately 8.45% in overall accuracy. However, it is still considered to be rather poor, with just up to 57.80% accuracy. Multi-date, single-polarised PALSAR images produced some improvement in classification performance in comparison with single-date SAR images – the classification accuracy using the SVM classifier increased by 8.77% and 8.47% for HH and HV polarised images, respectively. The SVM classifier gave 7.71% increase in accuracy for classifying multi-date, dual-polarised images. The improvement was 13.17% in the case of the ANN classifier. Uses of multi-date dual polarised data gave remarkable improvement in classification accuracy as compared with using multi-date single polarised datasets. The accuracy increased by 7.39% and 11.82% as compared with multi-

date HH images using SVM and ANN classifiers, respectively. The improvements were even more significant for the multi-date HV image with an increase of 11.51% (for SVM classifier) and 17.27% (for ANN classifier) in classification accuracy.

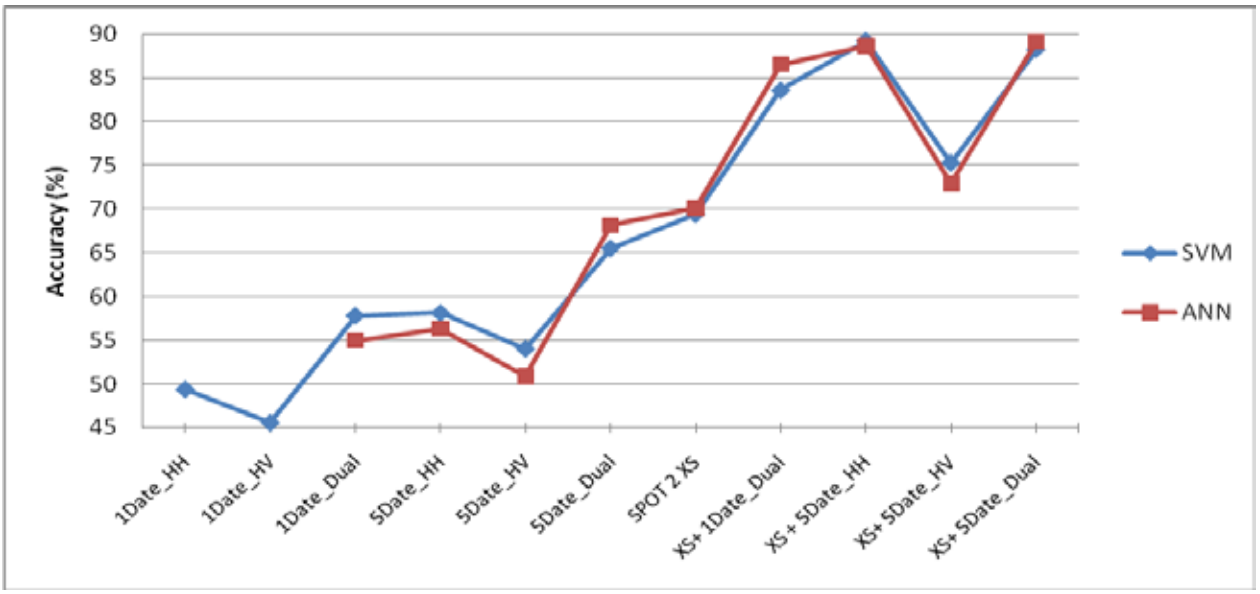


Fig 3: Land cover classification accuracy of different datasets

Sensitivity of like- and cross-polarisation to surface and volume scattering mechanisms were clearly demonstrated in the classification results. As can be seen in Figure 4 and Table 3, while the like-polarisation (HH) resulted in better accuracy for classifying urban structures and bare lands (which are subjected to surface scattering), cross-polarisation (HV) produced higher accuracy for both High Dense Vegetation (HDV) and Low Spare Vegetation (LSV) classes. The classification results of multi-date, dual-polarised PALSAR images (Table 3) highlighted a complimentary nature of like- and cross-polarised SAR data. Accuracy for most land cover classes increased significantly, except the Low Spare Vegetation class, compared to the case of classifying single-polarised data. These complimentary effects are also illustrated in Figure 4, which shows results of classification using single-date, single-polarized (HH or HV), and single-date, dual-polarised (HH+HV) SAR images.

Table 3. Producer and user accuracy (%) for SVM classifier applied to five-date PALSAR HH, HV, and five-date PALSAR dual-polarised (HH+HV) images

Land cover classes	Five-date HH		Five-date HV		Five-date HH+HV	
	Producer	User	Producer	User	Producer	User
Water Surface (WF)	40.89	76.44	47.48	76.84	47.80	85.89
Bare Land (BL)	57.73	47.91	51.21	55.59	58.15	59.72
Dense High Urban structures (DHU)	87.40	91.27	34.51	37.18	94.46	67.39
Low Flat Urban structures (LFU)	80.08	55.27	76.47	56.25	89.09	69.45
High Dense Vegetation (HDV)	48.88	48.43	54.19	52.95	57.16	65.84
Low Spared Vegetation (LSV)	37.96	41.26	57.87	46.82	51.04	47.83

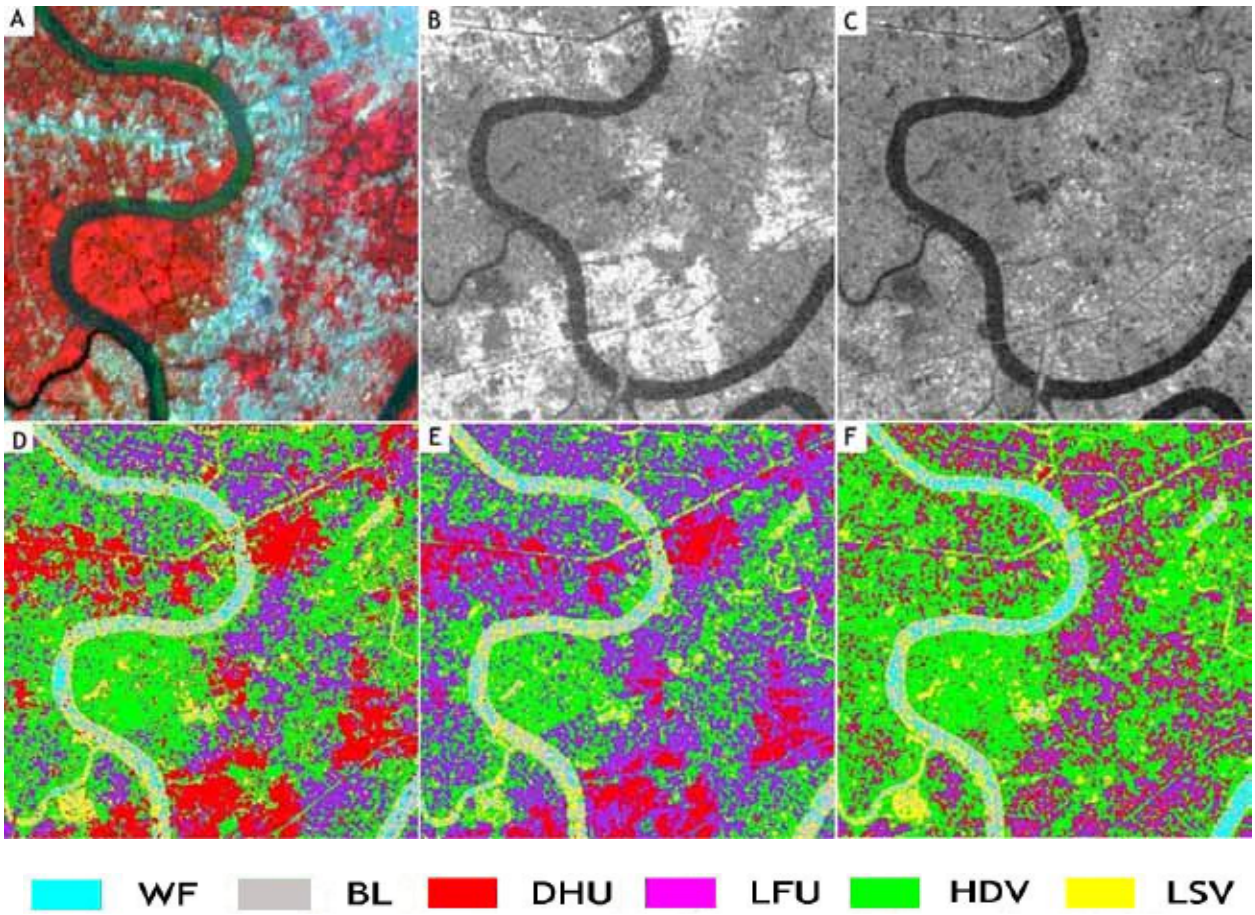


Fig. 4: Part of classification results using SVM classifiers on single-date, single-polarised (HH or HV), and single-date, dual-polarised HH+HV images: A) SPOT 2 multi-spectral false colour images, B) PALSAR HH polarised image, C) PALSAR HV polarised image, D) classification of single-date, dual-polarised (HH+HV) image, E) classification of single-date, single-polarised HH image, F) classification of single-date, single-polarised HV image

The SPOT multi-spectral image gave rather acceptable classification accuracy using the ANN and SVM classifiers. However, as can be seen in Figure 5and table 4, for SPOT 2 classification, the urban areas were confused with bare ground, and it is very hard to distinguish between Dense High Urban (DHU) and Low Flat Urban (LFU) structures due to their similar spectral properties. Consequently, the accuracy of these features, particularly DHU and LFU, were poor.

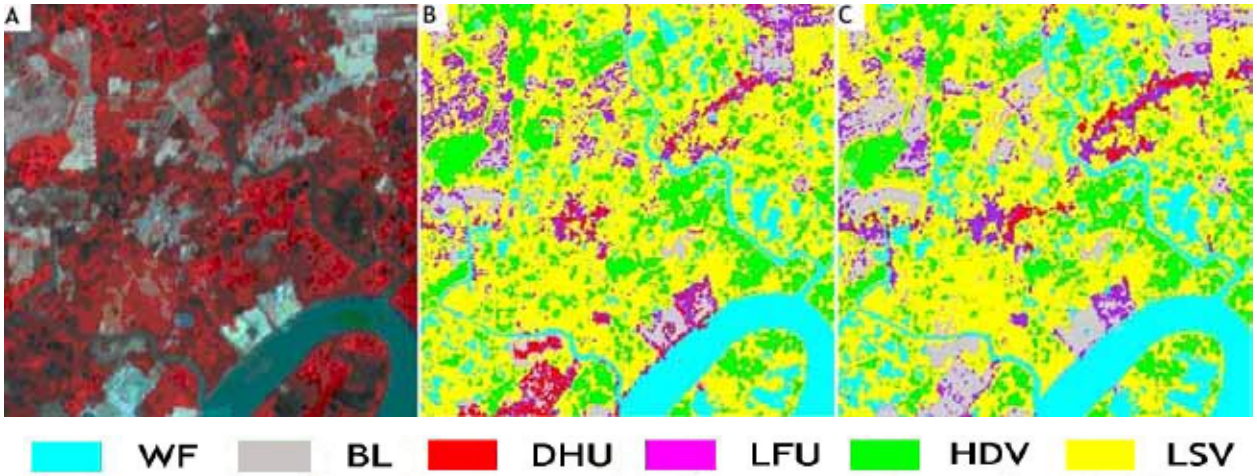


Fig. 5: Part of classifications results using SVM classifiers on a SPOT 2 multi-spectral image alone and the combination of SPOT 2 and multi-date PALSAR HH polarised images. A) the SPOT 2 multi-spectral image, B) classification of the SPOT 2 multi-spectral image, C) classification of a combination of SPOT 2 + five-date PALSAR HH polarised images

Table 4. Producer and user accuracy for ANN and SVM classifier applied to SPOT 2 multi-spectral images

Land cover classes	ANN		SVM	
	Producer accuracy(%)	User accuracy(%)	Producer accuracy(%)	User accuracy(%)
Water Surface (WF)	98.05	90.40	96.95	96.35
Bare Land (BL)	62.15	94.41	80.86	80.36
Dense High Urban structures (DHU)	2.47	56.76	44.17	38.11
Low Flat Urban structures (LFU)	86.05	46.47	26.94	35.68
High Dense Vegetation (HDV)	61.35	98.52	76.69	92.25
Low Spared Vegetation (LSV)	95.83	60.70	86.89	65.47

Table 5. Producer and user accuracy for ANN and SVM classifier applied to combination of SPOT 2 multi-spectral and five-date PALSAR HH polarised images

Land cover classes	ANN		SVM	
	Producer accuracy(%)	User accuracy(%)	Producer accuracy(%)	User accuracy(%)
Water Surface (WF)	90.81	96.63	94.31	99.66
Bare Land (BL)	96.32	92.81	98.32	94.83
Dense High Urban structures (DHU)	84.45	96.89	78.45	98.09
Low Flat Urban structures (LFU)	94.50	87.91	95.45	82.12
High Dense Vegetation (HDV)	74.34	90.88	74.13	95.39
Low Spared Vegetation (LSV)	90.51	70.45	92.36	71.76

The combination of SPOT 2 multi-spectral and multi-date SAR images gave the most significant increase in classification accuracy, using either ANN or SVM classifiers. As can be seen in Figure 5C and 6, table 5, confusion between Bare Land (BL), DHU and LFU was significantly reduced by integrating SPOT 2 and PALSAR images because of differences in the backscatter patterns of these features in SAR images. The highest overall accuracy was 89.27% for classifying the combined dataset of SPOT 2 + five-date PALSAR like- (HH) polarised images using the SVM classifier. The improvements compared to the case of using only SPOT 2 or five-date PALSAR HH images were 19.84% and 31.15%, respectively.

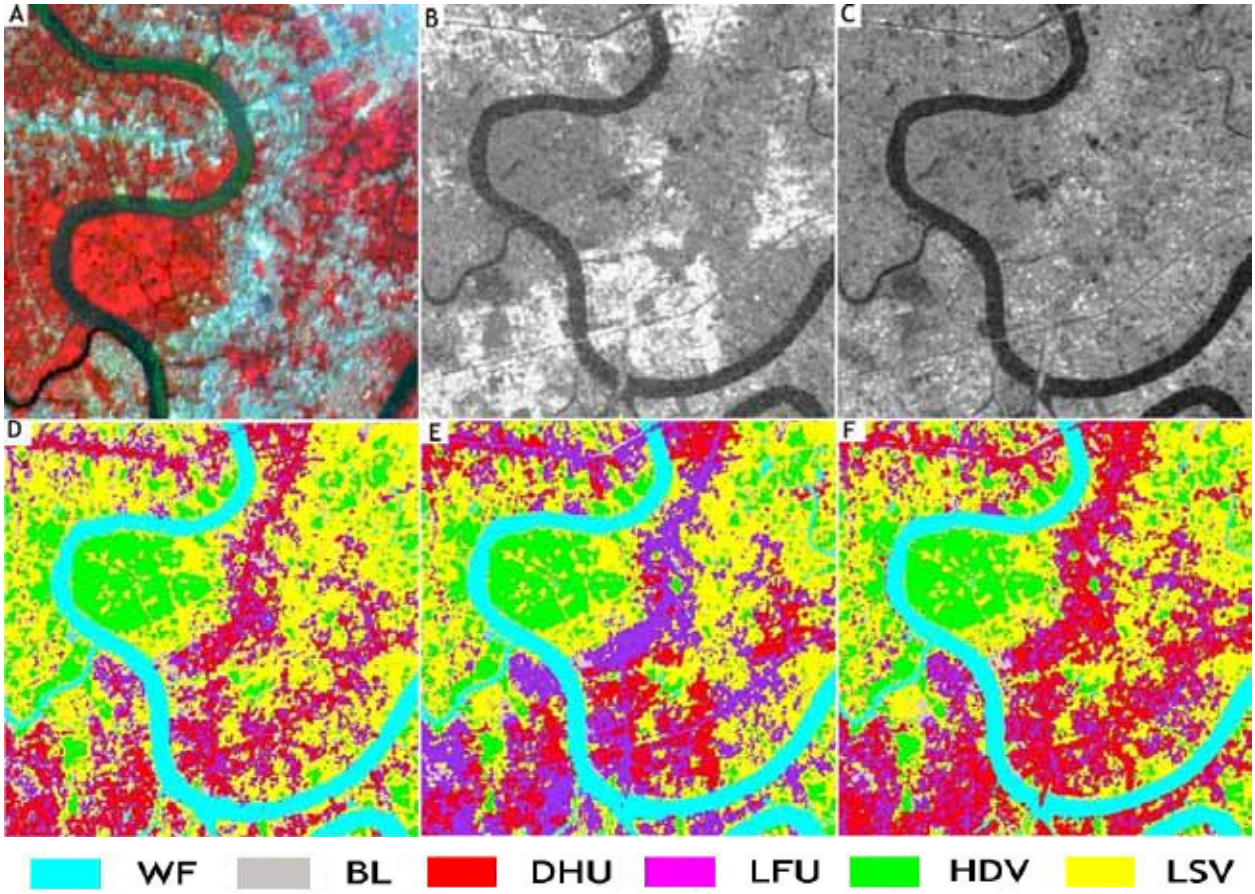


Fig. 6: Part of classifications results using SVM classifiers on a SPOT 2 multi-spectral images and combinating SPOT 2 and multi-date HH, HV dataset.
A) SPOT 2 multi-spectral false colour images, B) PALSAR HH polarised image, C) PALSAR HV polarised image, D) classification of a SPOT 2 multi-spectral images, E) classification of SPOT 2 + five-date PALSAR HH polarised image, F) classification of SPOT 2 + five-date PALSAR HV polarised images

Results of classification from best combinations of multi-date PALSAR images with textural data were given in the table 6. Sets of textures which gave the best classification accuracy for each classifier while incorporated with multi-dated PALSAR images are as follows:

Multi-date PALSAR datasets	Best incorporated textural feature	
	SVM	ANN
Five-date HH images	Homogeneity + Entropy Window size: 7x7	Variance + Entropy + Correlation Window size: 11x11
Five-date HV images	Variance + Entropy Window size: 11x11	Variance +Entropy + Correlation Window size: 9x9

The integration of multi-date single polarised SAR data with its selected best textured measures gave noticeable increase in the classification accuracy. While the combination of five-date cross-polarised images with its textures gave an increase of 4.29% and 7.51% in classification accuracy for ANN and SVM classifiers, improvements from the integration of five-date like-polarisation images with its textures were rather more significant, 6.95% and 8.91% for the SVM and ANN techniques, respectively. However, the contribution of these texture measures were neglected in a case of the five-date dual-polarised images, even it reduced the accuracy (0.41%) as for the ANN classifier.

Table 6. Comparison of land cover classification using multi-date PALSAR images including like, cross and dual polarised data and combination of these images with their best textural features.

Datasets	Accuracy (%)		Datasets	Accuracy (%)	
	SVM	ANN		SVM	ANN
Five-date HH	58.12	56.34	Five-date HH + Best textural features	65.07	65.25
Five-date HV images	54.00	50.89	Five-date HV + Best textural features	58.29	58.40
Five-date dual (HH+HV)	65.51	68.16	Five-date dual (HH+HV) + Best textural features	66.99	67.75

The results of classification indicated that the ANN and SVM algorithms were very effective for classifying multi-source remotely sensed data. There were no distinct differences in performance of these classifiers for most of the datasets tested. While the SVM classifier resulted in the highest overall classification accuracy (89.27%) for the combination of SPOT 2 multi-spectral and five-date PALSAR HH polarised datasets, the ANN classifier had the highest classification accuracy (89.16%) for the combined dataset of SPOT 2 multi-spectral and five-date PALSAR dual-polarised images.

Concluding remarks

The study reported here has demonstrated the advantages of the integrated approach, combining like- and cross-polarized SAR data with optical images for mapping land cover features using non-parametric classification algorithms.

Joint application of like- and cross-polarized (or dual-polarized) SAR images enhanced the performance of land cover mapping with respect to the case of using single-polarized images. The dual polarized dataset gave 65.51% and 68.16% of overall classification accuracy for SVM and ANN classifiers, respectively, while the figure were less than 59% for single polarized datasets.

The combination of optical multi-spectral data and either single- or dual-polarized, as well as multi-date, SAR images resulted in a significant improvement in classification accuracy up to 19.84% compared to use of only a SPOT 2 image and 31.15% compared to five-date single polarized dataset.

Incorporation of textural information with PALSAR images resulted in a considerable increase in classification accuracy, particularly for the like-polarized images with an improvement up to 8.91% .

The non-parametric classifiers SVM and ANN were successfully applied for classifying land cover features, particularly with complex datasets. The performances of these two classifiers were comparable.

Acknowledgement

The SPOT 2 multi-spectral image used for this study was granted by the National Remote Sensing Centre, The Ministry of Natural Resource and Environment of Vietnam.

ALOS PALSAR Level 1.1 product was processed by ERSDAC, Japan. Copyright of raw data belongs to METI and JAXA.

Reference

1. Argany, M., Amini, J., and Saradjian, M.R., "Artificial neural networks for improvement of classification accuracy in Landsat +ETM images", *Proceeding of Map Middle East Conference,* Dubai, March 26-29, 2006. Available at: http://www.gisdevelopment.net/proceeding/mapmiddleeast/2006/mm06pos_98.htm.

2. Chust, G., Ducrot, D., and Pretus, J., "Land cover discrimination potential of radar multitemporal series and optical multispectral images in a Mediterranean cultural landscape", *International Journal of Remote Sensing*, Vol. 25, No. 17, 2004, pp3513–3528.

3. Dixon, B., and Candade, N., "Multispectral landuse classification using neural networks and support vector machines: one or the other, or both", *International Journal of Remote Sensing*, Vol. 29, No. 4, 2008, pp1185–1206.

4. Erasmi, S., and Twele, A., "Regional land over mapping in the humid tropics using combined optical and SAR satellite data – a case study from Central Sulawesi, Indonesia", *International Journal of Remote Sensing*, Vol. 30, No. 10, pp2465-2478.

5. Herold, N.D., Haack, B.N., and Solomon, E., "Radar spatial considerations for land cover extraction", *International Journal of Remote Sensing*, Vol. 26, No. 7, 2005, pp1383-140.

6. Huang, C., Davis, L.S., and Townshend, J.R.G., "An assessment of support vector machines for land cover classification", *International Journal of Remote Sensing*, Vol. 23, No. 4, 2002, pp725–749.

7. Huang, H., Legarsky, J., and Othman, M., "Land-cover classification using Radarsat and Landsat imagery for St. Louis, Missouri", *Photogrammetric Engineering & Remote Sensing*, Vol. 73, No. 1, 2007, pp037–043.

8. Kavzoglu, T., and Mather, P.M., "The use of backpropagating artificial neural network in land cover classification", *International Journal of Remote Sensing*, Vol. 24, No. 23, 2003, pp4907-4938.

9. Kavzoglu, T., and Colkesen, I., "A kernel functions analysis for support vector machines for land cover classification", *International Journal of Applied Earth Observation and Geoinformation*, Vol. 11, 2009, pp352–359.

10. Kim, S.H., and Lee, K.S., "Integration of Landsat ETM+ and Radarsat SAR for land cover classification", *Trans Tech Publications*, Switzerland, 2005, doi:10.4028/www.scientific.net/KEM.277-279.838.

11. Lu, D., and Weng, Q., "A survey of image classification methods and techniques for improving classification performance", *International Journal of Remote Sensing*, Vol. 28, No. 5, 2007, pp823-870.
12. Mills, H., Cutller, M.E.J., and Fairbairn, D., "Artificial neural network for mapping regional-scale upland vegetation from high spatial resolution imagery", *International Journal of Remote Sensing*, Vol. 27, No. 11, 2006, pp 2177-2195.

13. Sheoran, A., Haack, B., and Sawaya, S., "Land cover/use classification using optical and quad polarization radar imagery", *Proceeding of ASPRS Annual Conference*, Baltimore, Maryland, March 9- 13, 2009. Available at: http://www.asprs.org/publications/proceeding/baltimore09/0089.pdf.

14. Waske, B., and Benediktsson, J.A., "Fusion of Support Vector Machines for classification of multisensor data", *IEEE Transactions on Geosciences and Remote Sensing*, Vol. 45, No. 12, 2007, pp4842-4845.

15. Waske, B., and Braun, M., "Classifier ensembles for land cover mapping using multitemporal SAR imagery", *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 64, Issue 5, 2009, pp 450-457.

Comparison of High Pass Filter and Wavelet Fusion Approaches for Fusion of Hyper-spectral and Multispectral Images

Yuanyuan Zhang [1], Linlin Ge [1] and Xiaojing Li [1]

[1] *Geodesy and Earth Observation System, School of Surveying and Spatial Information Systems, The University of New South Wales, Kensington, NSW 2052, Australia*

Summary: Image fusion is desired when both high spectral and spatial resolution are required. This paper focuses on fusion of hyper-spectral and multispectral images in order to have both high spatial and spectral resolutions on one image. Two fusion approaches, namely, high pass filter (HPF) and wavelet fusions, were applied. For the purpose of testing the methods CHRIS/Proba and Landsat 5 images were selected as the test data since they are available at no charge. The quality of both spatial resolution improvement and spectral preservation were assessed. Comparison of the correlation coefficient of fused image and Landsat TM image via high pass filter was the criteria of the spatial quality assessment. The basic statistics as well as the correlation coefficient of CHRIS image and fused image were used to evaluate the quality of spectral preservation. The result showed that HPF fusion approach performed better than wavelet fusion approach in terms of both spatial resolution improvement and spectral resolution preservation.

Keywords: image fusion, hyper-spectral image, multispectral image, HPF fusion, wavelet fusion

Introduction

Hyper-spectral remote sensor is sensitive to fine change of the spectral signature of the feature due to its high spectral resolution [1]. However, it usually has low spatial resolution which is difficult to detect the details such as the spatial distribution of the observations. In water quality monitoring, both high spatial and spectral resolutions are required in order to have a better understanding of spatial distribution and spectral details of the indicators [2]. According to [3], 3-20m spatial resolution is desired for the polluted zone monitoring while 20- 500m is desired for the detection of the distribution and motion of water pollution in lakes, rivers and bays. Therefore, the ideal sensor for inland water quality monitoring is supposed to have both high spatial and spectral resolution. However, in most cases, multispectral image has discrete spectral bands and relative high spatial resolution. On the contrary, hyper-spectral image can provide contiguous spectrum with relative low spatial resolution. [4] commended it is impossible to have both high spatial and spectral resolution in one sensor over a large area. The research done by [5] also showed that a sensor cannot obtain both in high resolution at one time. The reason is large IFOV is required in hyper-spectral sensor in order to acquire sufficient photon which maintains acceptable SNR to obtain the small bandwidth while small IFOV is required in multispectral sensor for obtaining broader spectral bandwidth with higher spatial resolution. The spatial resolution of multispectral sensor decreases as the number of bands increases [5]. Therefore, image fusion between multi-sensors becomes desirable. It can provide the better capability of

interpratation as well as more reliable results. This paper focuses on the image fusion of hyper-spectral and multispectral data. The aim is to enhance the spatial resolution of hyper-spectral image while maintain its spectral resolution. Landsat 5 multispectral and CHRIS/Proba hyper-spectral images were used in this study. The study is the preparation for future quantitative study on water quality monitoring through image fusion. High pass filter (HPF) fusion and wavelet fusion methods were tested in order to select one which is suitable for quantitative analysis in the further study. The rest of paper is organised as follows: the description of data used in the study is in section 2. Section 3 introduces the methodology used in the study. Section 4 discusses the results. Finally, conclusion and future work is given in section 5.

Data Used

CHRIS/Proba and Landsat 5 images were used in this study for testing the methods since they are available at no charge.

CHRIS/Proba

Proba with the Compact High Resolution Imaging Spectrometer (CHRIS) on board was launched on October 22nd 2001 as an Earth Observation mission [6]. CHRIS imagery can aquire image at five different angles (+55°, +36°, nadir, -36°and -55°) [7]. The figure below shows its acquisition geometry. In this study, only the image acquired at nadir angle was used.

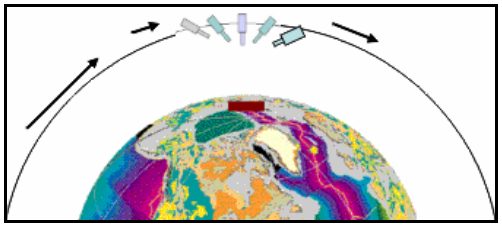


Fig. 1: CHRIS/PROBA acquisition geometry [8].

Table 1: The specification of CHRIS/Proba Mode 2 [6].

Date of launch	October 22th, 2001
Obit	Sun-synchronous
Sensor Altitude	615 km
Repeat interval	7 days
No. of bands	18 bands
Wavelength range	406-1036nm
Spectral resolution	8-11nm
Spatial resolution	17m

CHRIS/Proba operates 5 modes in terms of applications. Image of Mode 2 (water application) was used in the study. The specification shows in table 1.The image used was acquired on October 18th, 2005. The center location is 120.11E 31.29N (Lake-Taihu, China). The image below is CHRIS band 3 image of the study area.

Landsat 5

The Landsat program, managed by NASA is the longest earth observation mission and it continues to provide valuable images of the earth. It started in 1972. There are 7 satellites in the series. So far, only Landsat 5 and 7 satellites are in operation [9]. The image used in this study was acquired by Thematic Mapper (TM) sensor, so Landsat 5 is also called Landsat TM in the paper. The table below shows its specification.

Table 2: The specification of Landsat TM [9].

Date of launch	March 1st 1984	
Obit	Sun-synchronous	
Sensor Altitude	705km	
Repeat interval	16 days	
No. of bands	6 (band 1-5&7)	1 (thermal band 6)
Wavelength range	450- 2350 nm	10400-12500nm
Spectral resolution	70- 270nm	2200 nm
Spatial resolution	30m	120m

Landsat TM image used was acquired on October 17th 2005. Both images have no cloud cover. In addition, the one day difference between the Landsat and CHRIS acquisitions makes temporal decorrelation the least.

Methodology

Pre-processing

Several preparations must be carefully done before image fusion. Images from different sensors have different characteristics such as pixel size, image coverage, sensor response and dynamic range, sometimes even the orientation [10]. Therefore, corrections before fusion are vital to the quality of fused image. The image registration is extremely important as the pre-processing step. Many papers emphasise the essential importance of the data co-registration before processing image fusion [10-16]. Also there is a huge challenge in this issue. It is a primary factor to the spectral distortion in the fused image [10].

Landsat image has been georeferenced before it is available to the user. Image to image registration is applied on CHRIS image. In this way, CHRIS image can be georeferenced while completing image registration with Landsat TM.

Although CHRIS image has higher spatial resolution than Landsat TM and hence there seems to be no gain to fuse these two images; the purpose of this study is to choose the effective image fusion method for quantitative analysis for future studies when high resolution multispectral images are available. Therefore, for the purpose of testing the methods, the decrease of the spatial resolution was conducted on CHRIS. It would be twice lower than Landsat TM which is 60m. In this process, its spectral resolution would be remained.

Then CHRIS image has to be resampled to the same pixel size and coverage as Landsat image in order to continue fusion.

Image simulation

There are some spectral gaps in Landsat TM since it does not have continuous spectral bands as hyper-spectral image does. The purpose of image fusion in this study is to increase spatial resolution as well as preserve the spectral resolution of the hyper-spectral image; therefore, the missing bands which corresponded to CHRIS bands in Landsat TM have to be simulated. The linear interpolation method was used to simulate the missing data of Landsat TM. The equations are as follow.

$$\mathbf{M_i = W_1 * M_{i-1} + W_2 * M_{i+1}} \quad (1)$$

Where: M_i is the interpolated band; M_{i-1} and M_{i+1} are the bands that are before and after band M_i
W_1 and W_2 are the weights that are given to M_{i-1} and M_{i+1} according to the equation as follows:

$$\mathbf{W_1 = (\lambda_2 - \lambda_i) / (\lambda_2 - \lambda_1)} \quad (2)$$

$$\mathbf{W_2 = 1 - W_1} \quad (3)$$

Where: λ_1 and λ_2 are the wavelengths of M_{i-1} and M_{i+1}, λ_i is the wavelength of the interpolated band.

Image fusion

HPF fusion

The principle of HPF fusion is to extract the spatial information by filtering high spatial resolution image with high pass filter and add it with the low spatial resolution image. The purpose of the test was to extract spatial resolution from Landsat image and to add it to CHRIS data. HPF fusion method was one suitable approach to achieve the goal. To begin with, Landsat image was filtered with high pass filter with the filter kernel 5 to extract the spatial information. The filter kernel was determined by the ratio between Landsat TM and CHRIS image. Then, the standard derivation of filtered Landsat image and CHRIS image were calculated in preparation for adding the images at pixel level. Finally, adding the images pixel by pixel according to the formulas:

$$\mathbf{PF = PH + (PFM * W)} \quad (4)$$

Where: PF is fused pixel
PH is hyper-spectral pixel
PFM is the filtered Multispectral pixel
W is the weighting multiplier for HPF image value which is determined using the equation as follows:

$$\mathbf{W = \left(\frac{SD(HI)}{SD(FM)} x\ M\right)} \quad (5)$$

Where: SD(HI) is the standard deviation of the hyper-spectral image to which the filtered multispectral image is being added.
SD(FM) is standard deviation of the HPF image.
M is the modulating factor to determine the crispness of the output image. In this case M is 0.25.

Wavelet fusion

Wavelet fusion method was developed from the signal processing concept [17]. The wavelet method uses the low frequency component in low spatial resolution image replace the low frequency component in high spatial resolution image. The principle of wavelet fusion method is similar to HPF approach which was to obtain spatial detail of Landsat image and fuse it with CHRIS data. Remote sensing images are two-dimensional digital signals. Three detailed components (vertical, horizontal and diagonal) of Landsat TM were decomposed through wavelet filter. Then CHRIS image was decomposed into three detailed components (vertical, horizontal and diagonal) via wavelet filter. Finally, the approximation component of Landsat TM was fused to the corresponding approximation component of CHRIS image.

Quality evaluation method

Fused image is supposed to provide increased interpretation capabilities and reliable results since both advantages of multispectral and hyper-spectral images are available on the fused image. The criteria were based on the improvement of spatial resolution and the preservation of spectral characteristics. To begin with, visual comparison was conducted. It is the most direct way to tell the difference. Then correlation coefficient, minimum, maximum, mean, mode value and standard deviation are used as the criteria to assess spatial and spectral quality of the fused image.

Spatial quality evaluation

Comparison of the correlation coefficient of the fused image and Landsat TM image via high pass filter was the criteria of the spatial quality assessment in this study [20]. High pass filter can extract the high frequency information which is related to spatial resolution. The value range is supposed to be between -1 and 1; however, since the fused image was generated based on Landsat TM, the value range should be from 0 to 1. The higher the spatial correlation between the fused image and Landsat TM is, the higher the value of correlation coefficient is.

Spectral quality evaluation

Some comparisons between the fused image and CHRIS image were used to evaluate the spectral consistence. The basic statistics of image like minimum, maximum, mean, mode value and standard deviation were taken into account for spectral consistence assessment. Smaller spectral distortion has less difference of basic statistics between fused and original images. In addition, spectral correlation between the fused and CHRIS images was another criteria of assessing the fused image. The higher correlation means the better spectral preservation. Finally, a visual appearance of complete spectrums of fused and CHRIS images was presented to compare the difference.

Results and discussion

Due to the word limit of the paper, the true color fused image composed by CHRIS image band 3, 8 and 12 and corresponded Landsat bands was used in the next section for quality assessment.

Results of fusion

Fig. 2 below shows the comparison of CHRIS image and fused image using two methods. It can be seen visually both approaches improve the spatial resolution of the original image.

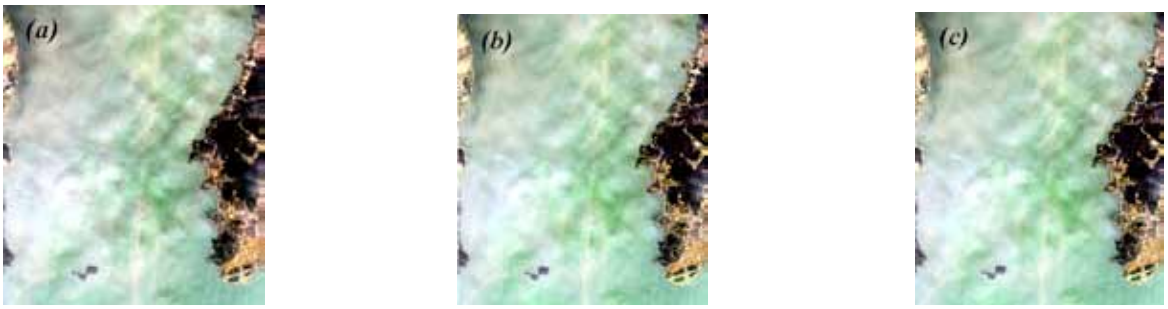


Fig. 2: (a) CHRIS image before fusion; (b) HPF fused image and (c) wavelet fused image.

Quality assessment

Spatial quality assessment

The figures below (fig. 3-5) show the spatial correlation between Landsat TM and the fused images. The figures show the HPF fused image had higher spatial correlation with Landsat TM which was around 0.6 to 0.83 than wavelet fused image which was around 0.1 to 0.3. It indicates that HPF fusion had better ability than wavelet fusion on extract the spatial information. This may because HPF process could extract the spatial information better. When comparing three charts, it can be seen that the spatial correlation between Landsat TM and fused band 3 image was the highest while the fused image band 12 was the lowest. From the spatial correlation charts, it can be seen that HPF fusion performed much better than wavelet fusion in spatial resolution improvement.

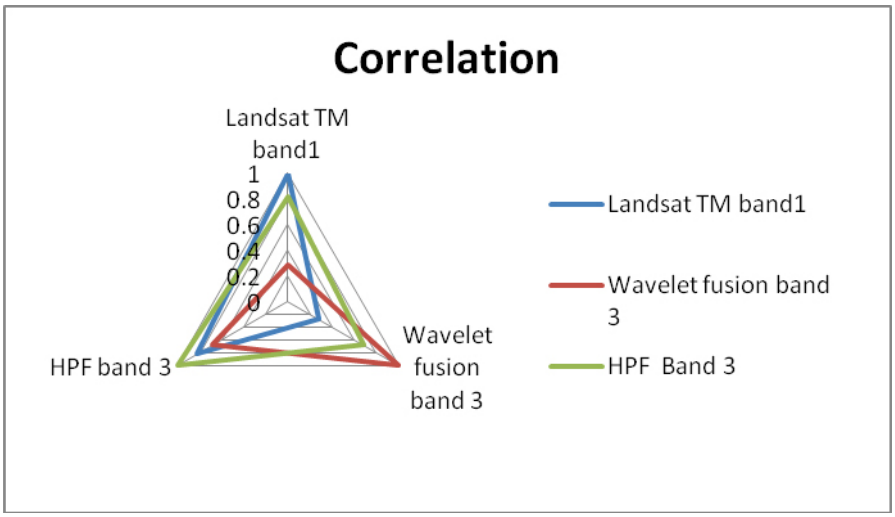


Fig. 3: Correlation of Landsat TM (band 1) and fused images (band 3)

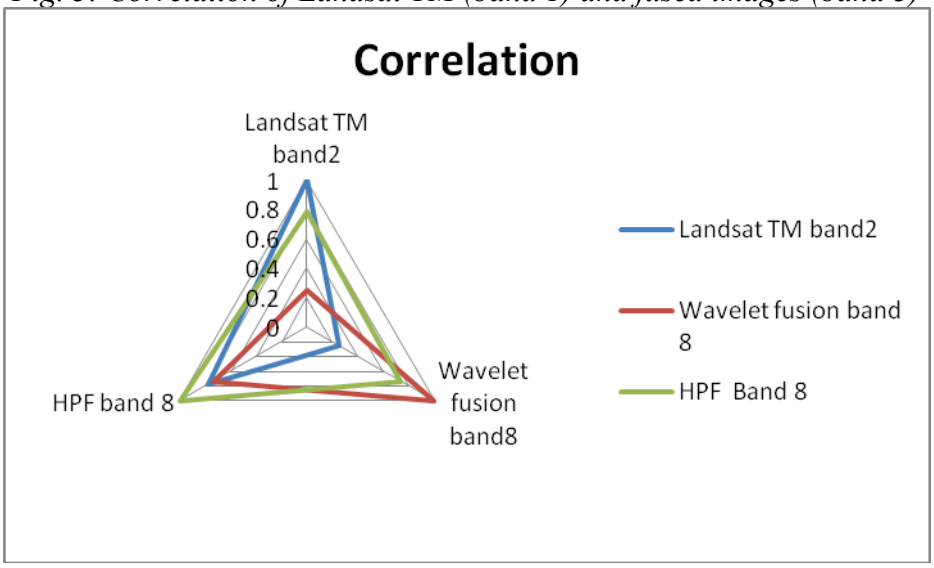


Fig. 4: Correlation of Landsat TM (band 2) and fused images (band 8)

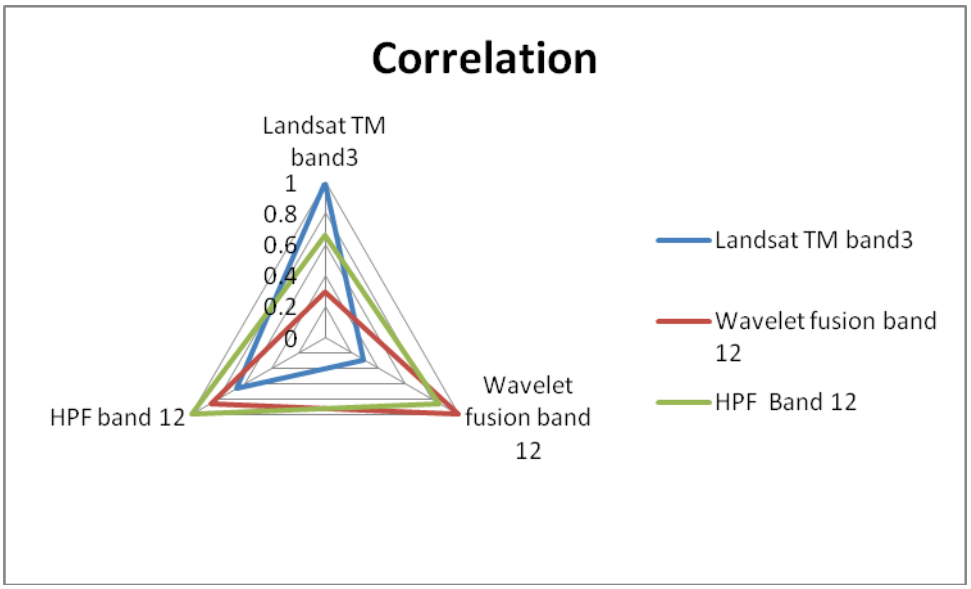


Fig. 5: Correlation of Landsat TM (band 3) and fused images (band12)

Spectral quality assessment

The tables below illustrate the basic statistics of CHRIS and the fused images using two methods respectively. Comparing all the variables, HPF fused image was more consistent with the original one than the wavelet fused image.

Table 3: Basic statistics of CHRIS and the fused band 3 images

Basic Stats	Stdev	Min	Mean	Mode	Max
CHRIS band 3 image	0.070374	0.742417	1.006742	1.027323	1.4992
Wavelet fused band 3 image	0.074232	1.30619	1.609253	1.620748	2.028824
HPF fused band 3 image	0.072988	0.698424	1.006738	1.022934	1.671955

Table 4: Basic statistics of CHRIS and the fused band 8 images

Basic Stats	Stdev	Min	Mean	Mode	Max
CHRIS band 8 image	0.102244	0.61664	1.012464	1.03589	1.817862
Wavelet fused band 8 image	0.138193	1.161697	1.685141	1.702547	2.404191
HPF fused band 8 image	0.106604	0.543644	1.01246	1.052212	2.185225

Table 5: Basic statistics of CHRIS and the fused band 12 images

Basic Stats	Stdev	Min	Mean	Mode	Max
CHRIS band 12 image	0.116245	0.570717	1.005848	1.021805	2.34037
Wavelet fused band 12 image	0.12319	0.976151	1.526527	1.553544	2.612097
HPF fused band 12 image	0.122542	0.502172	1.00584	1.022098	2.828156

Comparing the spectral correlation results, both methods obtained the best result in band 8 while the worst in band 12. Comparing the performance of HPF fusion on the 3 bands, it had the best performance in band 8 while the worst in band 12. On the other hand, wavelet fusion worked the best in band 8 and the worst in band 3.

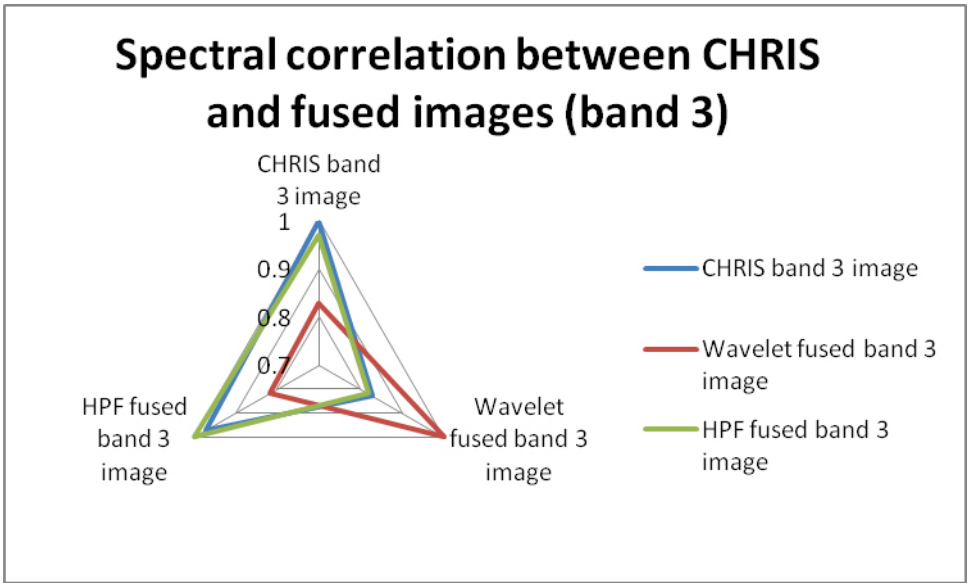


Fig. 6: Correlation of CHRIS image and fused images (band 3)

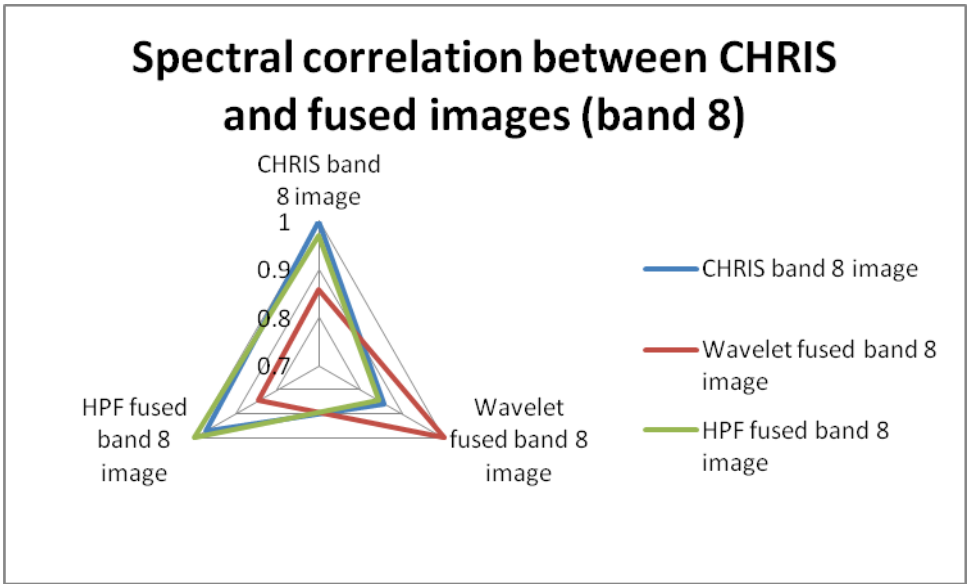


Fig. 7: Spectral correlation of CHRIS image and fused images (band 8)

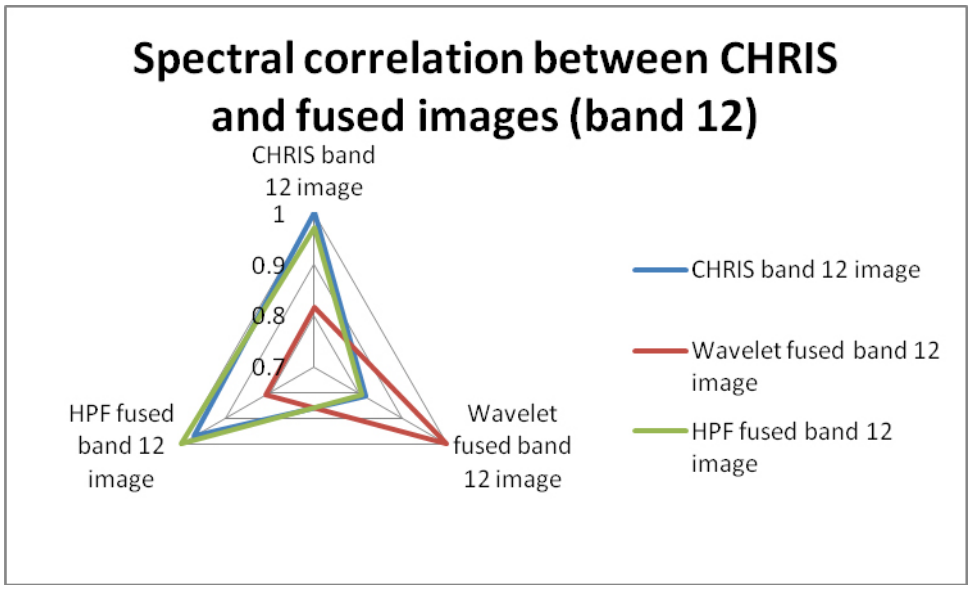


Fig. 8: Correlation of CHRIS image and fused images (band 12)

The plot below shows the spectral profile of CHRIS and fused images. CHRIS image and HPF fused image have almost the identical spectrum. The spectrum of wavelet fused image can even not compare with that of the original image.

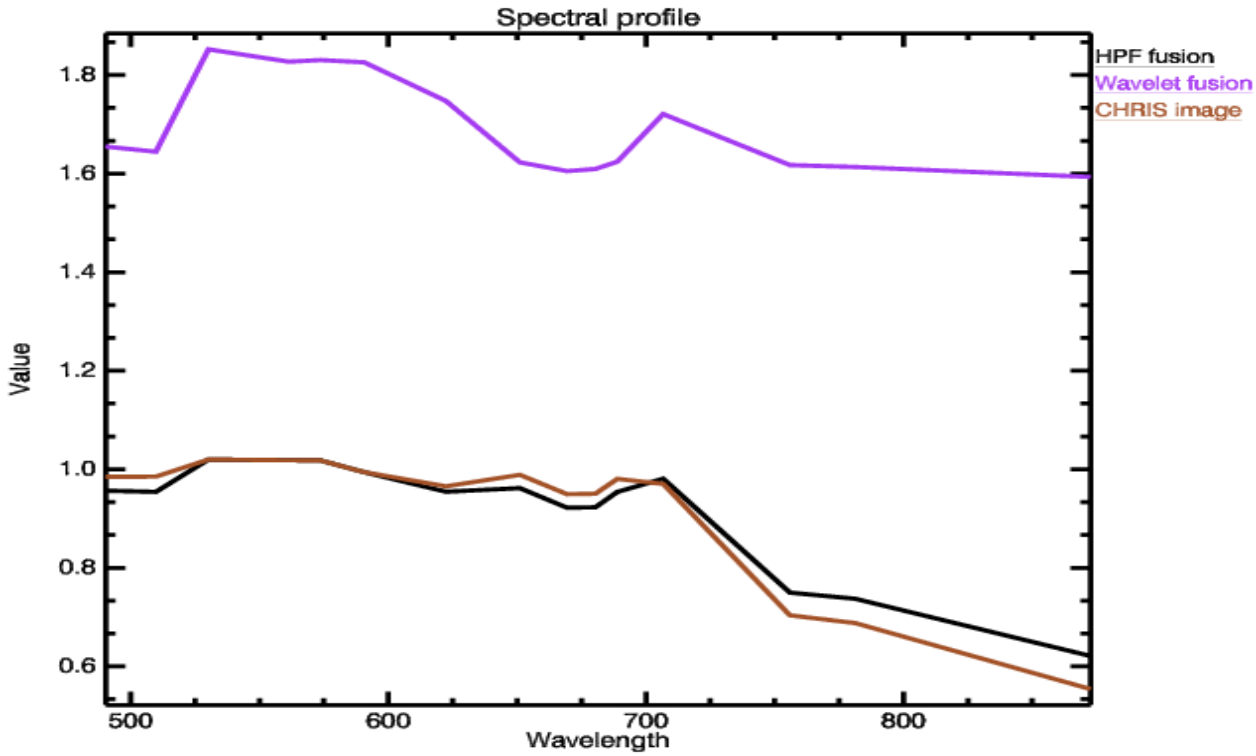


Fig.9: Spectrum of CHRIS image and fused images

The qualified fused image should have the high spectral consistence with CHRIS image and minimum spatial difference with Landsat TM. Therefore, all the statistics showed HPF fusion

method performed better than wavelet fusion method in terms of spatial resolution improvement and spectral reservation.

Conclusion and future work

The results show both wavelet and HPF fusion methods can improve the spatial resolution of lowered CHRIS image. However, HPF fusion performs better than wavelet fusion. In addition, HPF fusion performs much better than wavelet fusion in terms of spectral preservation of CHRIS image. Therefore, HPF fused image has the potential in quantitative analysis. Since the work in this paper is the preparation of the quantitative analysis of water quality in the future, HPF fused image is going to be used for retrieval of water quality parameters. The aim of using fused image on water quality monitoring is to improve the accuracy since the higher spatial resolution is achieved and spectral resolution well preserved.

Acknowledgements

We thank ESA for providing CHRIS/Proba image and NASA for providing Landsat TM image to this study.

References

1. Wen, J. G., Xiao, Q., Yang, Y. P., Liu, Q. H. & Zhou, Y., "Remote Sensing Estimation of Aquatic Chlorophyll-a Concentration Based on Hyperion Data in Lake Taihu", *Journal of Lake Sciences,* Vol. 18, No. 4, 2006, pp. 327-336.
2. Li, H., Manjunath, B. S. and Mitra, S. K. "Multi-sensor image fusion using the wavelet transform", *Proceedings of the 1994 IEEE Imternational Conference*, Vol. 1, 1994, pp. 51-55.
3. Tong, X. H., Xie, H., Zhang, J., Zhang, Y. L., Zhao, J. F. and Qiu, Y. L., "Multi-spectral remote sensing based water quality monitoring for Lake Tai", *MIPPR 2005: SAR and Multispectral Image processing,* 2005.
4. Ma, Y. H., "Image Fusion from High Spatial Resolution Remote Sensing and High Spectral Resolution Remote Sensing", *Infrared*, 2003, pp. 11-16.
5. Jensen, J. R., *Remote sensing of the environment : An earth resource perspective*, New Jersey: Prentice-Hall, 2000.
6. ESA, Earth Observation Missions *[online,]* Available at: http://earth.esa.int/missions/thirdpartymission/proba.html [11 September, 2010], 2010.
7. Begiebing, S., Bach, H., Waldmann, D., and Mauser, W., "Analyses of spaceborne hyperspectral and directional CHRIS data to deliver crop status for precision agriculture", *[online],* Available: http://www.preagro.de/Veroeff/ECPA05_VISTA.pdf, [11 September, 2010], 2005.
8. Sugianto, Merton, R. N. and Laffan, S., "An overview of the CHRIS/PROBA Mission: A new generation of multi-angle hyperspectral remote sensing and its application to agriculture", *Proceedings of 3rd FIG Conference, Jakarta*, 3 - 7 October, Indonesia, 2004.
9. NASA, Landsat 5 *[online,]* Available at: http://landsat.gsfc.nasa.gov/about/L5_td.html [11 September, 2010], 2010.
10. Jazaeri, A., Enhancing hyperspectral spatial resolution using multispectral image fusion: A wavelet approach, United States -- Virginia, George Mason University, Ph.D., pp. 189, 2007.

11. Ehlers, M., Jadkowski, M. A., Howard, R. R. and Brostuen, D. E., "Application of SPOT Data for Regional Growth Analysis and Local Planning", *Photogrammetric Engineering and remote Sensing,* Vol. 56, No. 2, 1990, pp. 175-180.
12. Ehlers, M. and Klonus, S., "Quality assessment for multitemporal and multi sensor image fusion", *Remote Sensing for Environmental Monitoring, GIS Applications, and Geology VIII*, September 15, 2008 - September 18, 2008, Cardiff, Wales, United kingdom, SPIE, 2008.
13. Huang, Z. and Yu, X., "Application of several non-negative matrix factorization-based methods in remote sensing image fusion", *5th International Conference on Fuzzy Systems and Knowledge Discovery*, FSKD 2008, October 18, 2008 - October 20, 2008, Jinan, Shandong, China, Inst. of Elec. and Elec. Eng. Computer Society, 2008.
14. Ghassemian, H., "Multi-sensor remote sensing image fusion based on Retina-Inspired model", *2009 IEEE Symposium on Industrial Electronics and Applications*, ISIEA 2009, October 4, 2009 - October 6, 2009, Kuala Lumpur, Malaysia, IEEE Computer Society, 2009.
15. Metwalli, M. R., Nasr, A. H., Osama S. N. and El-Rabaie, S., "Image fusion based on principal component analysis and high-pass filter", *2009 International Conference on Computer Engineering and Systems*, ICCES'09, December 14, 2009 - December 16, 2009, Cairo, Egypt, IEEE Computer Society, 2009.
16. Pande, H., Tiwari, P. S. and Dobhal, S., "Analyzing hyper-spectral and multi-spectral data fusion in spectral domain", *Journal of the Indian Society of Remote Sensing,* Vol. 37, No. 3, 2009, pp. 395-408.
17. Mallat, S. G., "A theory for multiresolution signal decomposition: then wavelet representation", *IEEE Trans. Pattern Anal*. Mach. Intell. Vol. 11, 1989, pp. 674-693.
18. Darvishi, B. A., Kappas, M. and Erasmi, S. "Hyper-Spectral/High-Resolution Data fusion: Assessing the Quality of EO1-Hyperion/Spot-Pan & Quickbird-MS Fused Images in Spectral Domain", 2000.
19. Hsu, S. M. a. B., H. K., "Multisensor Fusion with Hyperspectral Imaging Data:Detection and Classification", *Lincoln Laboratory Journal,* Vol.14**,** No. 1, 2003, pp145-159.
20. Stathaki, T. ed. *Image Fusion: Algorithms and Applications*, Academic Press, 2008.

Monitoring Geologic Carbon Sequestration with Radar Remote Sensing

Rattanasuda Cholathat[1], Linlin Ge[1], and Xiaojing Li[1]
[1] *Geodesy and Earth Observing Systems Group (GEOS),*
School of Surveying & Spatial Information Systems,
University of New South Wales (UNSW), Sydney, NSW 2052, Australia

Summary: Geologic Carbon Sequestration becomes an essential instrument which has been aimed for mitigation of global warming. More understanding of storage processes and assessing environmental, safety, and health impacts in the event of a leak to the atmosphere are certainly important. Therefore, monitoring is one of the answers enabling technologies for CO_2 storage. It is expected to support multiple purposes from providing information about safety and environmental concerns to optimize storage projects, deal with unintended leakage and address regulatory, legal and social issues. Particularly, radar remote sensing will be a necessary tool to monitoring subsurface deformation after injecting CO_2 underground. This paper will discuss the use of a cost-effective technique, DInSAR for the first geosequestration project in Australia. It can contribute to measuring ground deformation to assess the movement of CO_2 in the ground-surface up to centimeter scale.

Keywords: Geologic Carbon Sequestration, CO_2 migration, DInSAR, monitoring GCS

I. Introduction

In order to maintain the atmospheric concentration of CO_2, there are some schemes which aim to mitigate global warming including increasing energy efficiency or by switching from coal to renewable energy, fossil fuel decarbonization technology and carbon capture sequestration, using nuclear energy (Canada global warming, 2009). However, the global commercialization and implementation of renewable energy sources has proven technologically difficult and slow moving. There is a gap between renewable energy production and energy demand, which for several decades will be filled by fossil fuels. Carbon Capture and Storage (CCS) will allow the continued use of fossil fuels, while the negative environmental impact is minimized. Scenarios from the International Energy Agency (IEA) also indicate that the potential for reduced CO_2 emissions through enhanced energy efficiency and increased renewable energy production is limited (IEA, 2008).

As a result, Geologic Carbon Sequestration is an essential instrument to remove massive amounts of CO_2 before it is released into the atmosphere. It is very important that Geologic Carbon Sequestration is supplementary to the continued effort of increasing use of renewable energy as well as energy efficiency measures. In addition, these technologies are capable of deploying widely across the globe in many different economic sectors and in many different locations. Moreover, CCS has the potential to be an essential technology to significantly reduce

greenhouse gas emissions and allow the continued use of fossil fuels for energy security, without damaging climate security.

What is Geological Sequestration

Carbon capture sequestration or Carbon Capture and Storage (CCS) is the process of capturing carbon dioxide from emission sources such as power plants, industrial facilities, and the transport and permanent disposal into geological formations deep underground (IEA, 2008). *Geologic Carbon Sequestration (GCS)* is normally taken to mean geological storage which represents the CO_2 injected deep underground using the most advanced technology offering large, long term storage capacity and minimal environmental effects.

In figure 1, there are three components of the scheme of carbon capture and storage which involves:

- Capture of CO_2 either before or after combustion of the fuel
- Transport of the captured CO_2 to the site of storage, and
- Injection and storage of the CO_2.

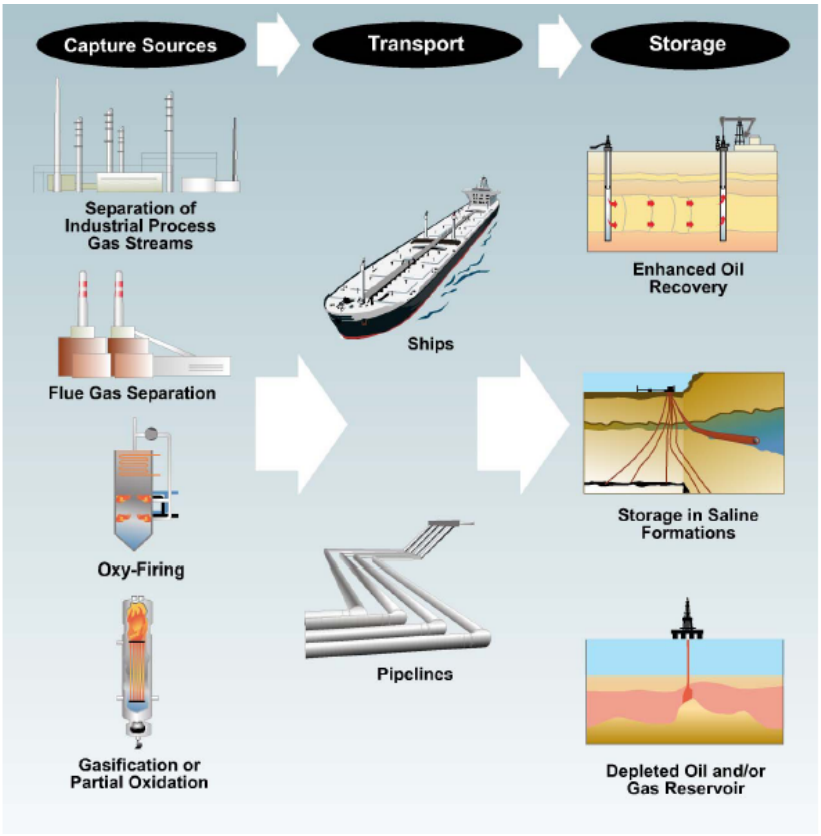


Figure 1: CCS Process Chain (Adapted from Energy Technology Analysis, IEA; prospects from CO_2 capture and storage 2004)

There are three generic process routes for capturing CO_2 from fossil fuel combustion plants:

- Post-combustion capture
- Pre-combustion capture.
- Oxy-fuel combustion.

Each of these processes involves the separation of CO_2 at high pressure, so that it can readily be transported to a storage site. The next step is to inject CO_2 in deep geological formations. This process involves many of the same technologies already used by oil and gas exploration and production industry as shown in figure 2. For instance, well-drilling technology, injection technology, computer simulation of storage reservoir dynamics and monitoring methods from existing applications are being developed further for design and operation of geological storage. According to The Intergovernmental Panel on Climate Change (IPCC), there are three types of geological formations that have received extensive consideration for the geological storage of CO_2 including oil and gas reservoirs, deep saline formations and unmineable coal seams (2005).

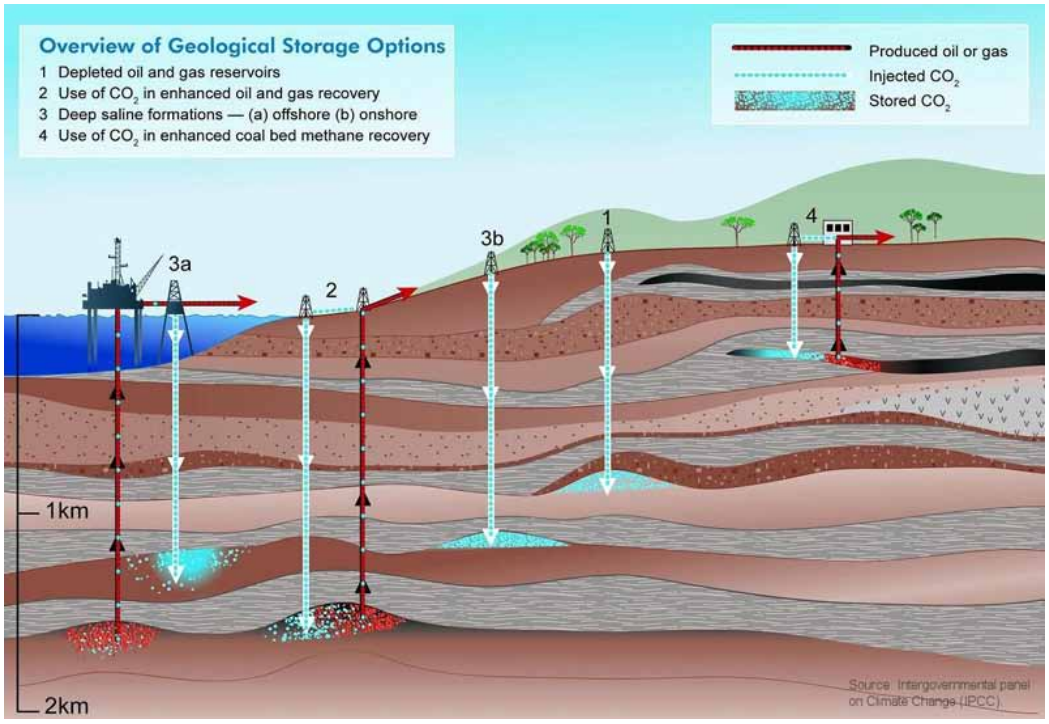


Figure 2: Overview of Geological Storage Option (IPCC, 2005)

Monitoring geological sequestration

According to The US National Energy Technology Laboratory (2008), the monitoring activities will vary among phases of geologic storage project cycle which involve four phases. Firstly, *pre-operation phase* is carried out the characteristic of geology, baseline condition such as time scale and project risks. Secondly, *operation phases* represent the period of time during which CO_2 is injected into the storage reservoir. Thirdly, *closure phase* refers to the period after injection has stopped, during which wells are abandoned and plugged. All equipments and facilities are removed; only necessary monitoring equipment is retained. Lastly, *post-closure* phase during

which ongoing monitoring is used to demonstrate that the storage project is performing as expected until it is safe to discontinue further monitoring. If it is satisfactorily demonstrated that the site is stable, monitoring will no longer be required except in the very unlikely event of leakage (NETL, 2008). Besides, at a properly managed CCS facility, the chance of perceptible CO_2 leakage from underground storage formation is very small. The fundamental objective for the measurement, monitoring, and verification of stored CO_2 is to demonstrate the long-term retention of stored CO_2 to regulators and the public, and to assist modeling of CO2 migration underground.

On the other hand, new and improved measurement and monitoring techniques and standards for their use need to be developed to provide proof of public and environmental safety and of each CCS project's effectiveness in mitigating climate change. Moreover, monitoring is one of the answers enabling technologies for CO_2 storage. It is expected to support multiple purposes from providing information about safety and environmental concerns such as used to optimize storage projects, deal with unintended leakage and address regulatory, legal and social issues. Harris et al. (2009) state that there are two important reasons for monitoring geological sequestration which are assessing the efficiency of the injection process and to verify safe containment. Surface deformation monitoring has become an important tool for observation of underground fluid flow. Harris also mentions that displacement of the Earth's surface will occur when pore pressure changes in the subsurface (2009). In a closed reservoir, the pressure changes would be caused by the injection of CO_2. This deformation can be monitored using satellite radar interferometry (InSAR), ground-based GPS, or surface tiltmeters, depending on the magnitude of the deformation and surface conditions.

II. Methodology

More understanding of monitoring CO_2 will provide the effectiveness of assessing environment, safety, and health impacts in the event of a leak to the atmosphere are certainly important. Therefore, several measures will be necessary to facilitate the successful adoption of CCS technology, particularly satellite radar remote sensing.

In Figure 3 shows that Synthetic Aperture Radar (SAR) is a combination of signal-processing techniques with satellite orbit information to create an interferogram pattern. Both amplitude and phase information are retained from the returning echoes. An interferogram is a phase difference image of two SAR-images acquired at slightly different positions. When the two images are not taken simultaneously the technique is called repeat-pass interferometry. An interferograms can be converted to a digital elevation model (DEM) with meter accuracy (Ferretti et al, 2007).

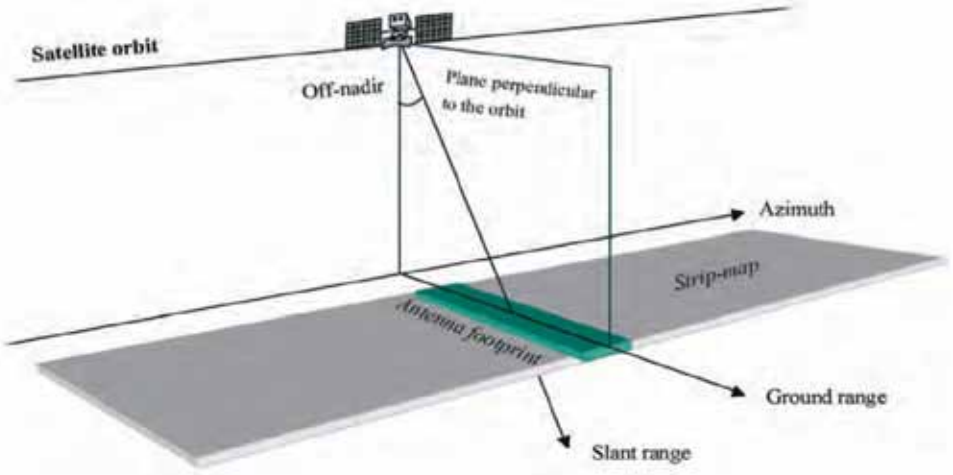


Figure 3: SAR system from the satellite (Ferretti et al, 2007)

Differential Interferometry for Synthetic Aperture Radar (DInSAR) technology is also a space borne measurement method or another explanation: a satellite based remote sensing technique which is able to detect ground motion with millimetric resolution using radar satellite images (Ge et al, 2004, 2007). The goal of using DInSAR is to separate phase contributions due to topography and displacement in order to get the deformation field. A digital elevation model or another SAR-image pair is needed to get the topographic phase, which will be subtracted from the deformation interferogram (Ferretti et al, 2007). Therefore, the distance to a specific location on the Earth's surface can be measured with impressive accuracy by exploiting a sequence of satellite images. It is possible to monitor height changes versus time. The phase of the differential interferogram $\mathbf{\Phi}$ is referred as

$$\mathbf{\Phi} \approx \varphi_{disp} + \varphi_{topo} + \varphi_{atmo} + \varphi_{noise} \qquad [1]$$

Where: disp = surface displacement
topo = topographic information
atmo = atmospheric delay
noise = phase noise

The DInSAR process is shown in figure 4;

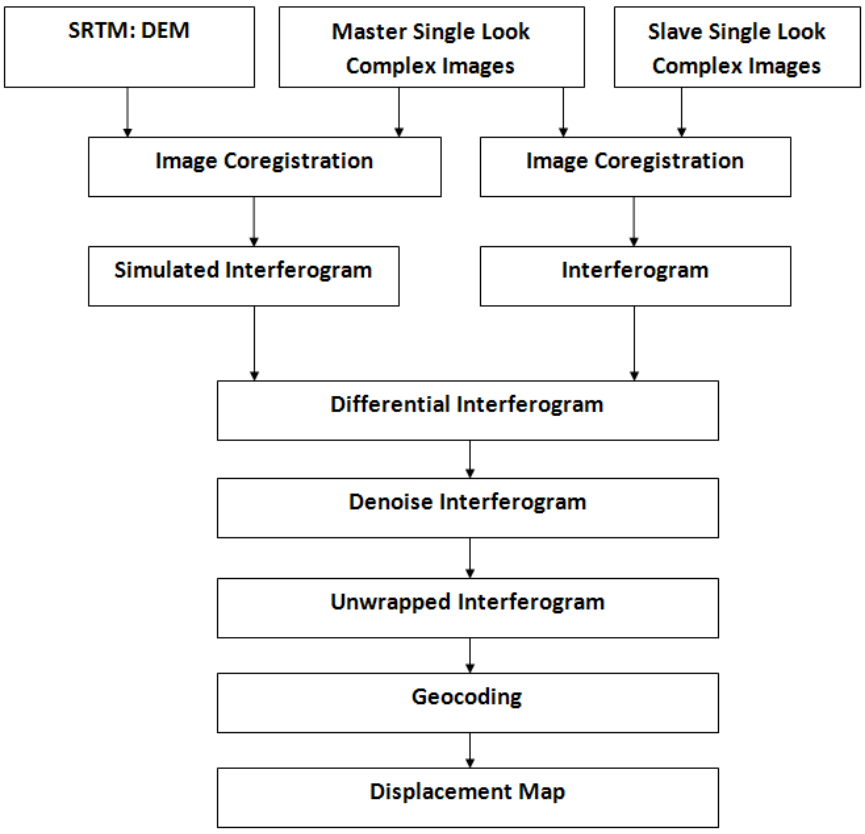


Figure 4: DInSAR process Flowchart

For CO_2 sequestration, there are several of key technologies to monitor the injection and the subsurface movement and storage of CO_2. A time lapse satellite image (using DInSAR technology) is the one which can measure ground deformation to assess the movement of CO_2 in the subsurface Mathioson (2009). DInSAR can contribute to an evaluation of the storage stability, since millimetric surface uplift could be one of the indicators for a potential storage leak. By calculating the phase differences for a signal received from the same location for different calendar times, it is additionally possible to measure the relative changes with even higher accuracy. To illustrate, Onuma and Ohkawa (2008) notice that the surface deformation around three CO_2 injections well at In Salah, Algeria was detected and analyzed by satellite borne SAR data. The result of this survey showed that the surfaces heave was at a rate of up to 7 mm/ year. Moreover, Vasco et al. (2008) apply a set of interferometric synthetic aperture radar (InSAR) range change data associated with the injection of CO_2 by using deformation in the material overlying to monitor pressure change at depth in active reservoir. This survey mentions that geodetic data, such as leveling measurements, in order to qualify surface deformation relate to fluid production or injection. Truly, the techniques of this study is to use InSAR to detect deformation within the reservoir and influence displacement within the surrounding medium including producing measurable deformation at the earth's surface.

III. Case study: Otway, Australia

The CO2CRC Otway storage Project (figure 5) is the first advanced geosequestration project in Australia and the world's largest research and demonstration project. As figure 6 shows the project which began in April 2008 had injected about 100,000 tons of CO_2 through a depleted gas field at a depth of 2 kilometers for a two year period (Cook, 2000).

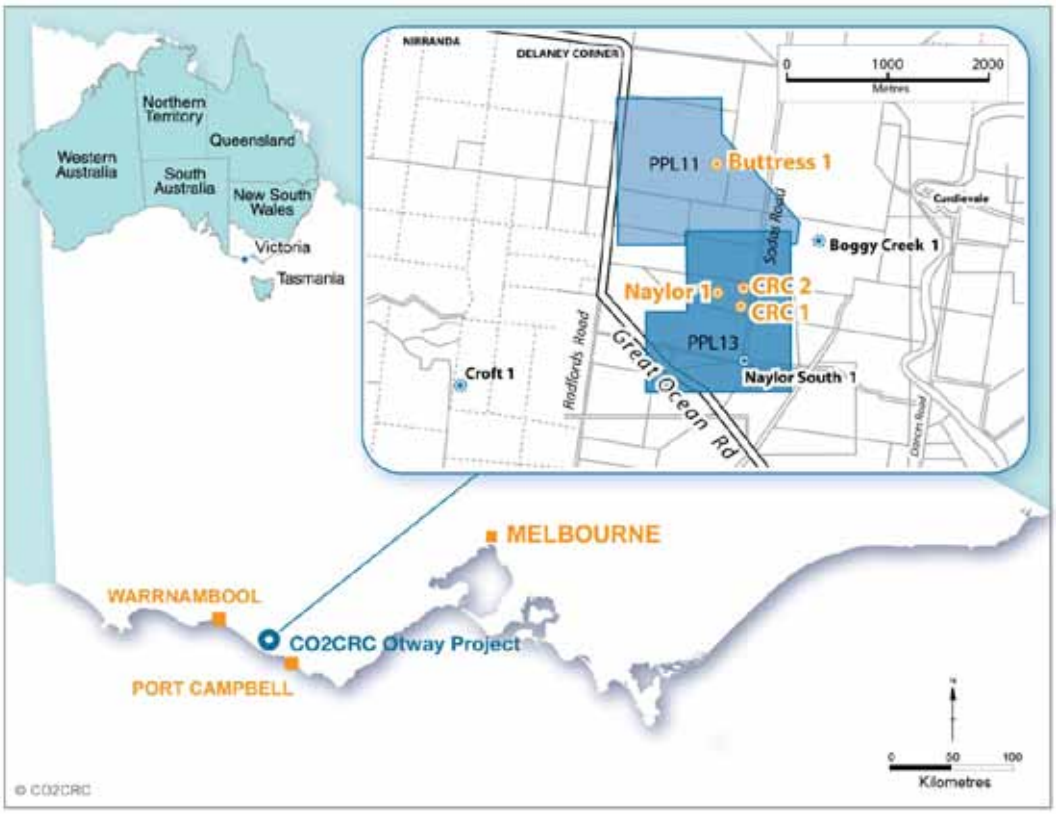


Figure 5: CCS projects in Australia, Right: Otway CO2CRC project (CO2CRC, 2011)

Figure 6: Injection site for the Otway project in South Western Victoria (CO2CRC, 2009)

There are multiple tasks of monitoring and verification procedures for this project which aim at effectiveness and safety of the site for CO_2 storage, in other words assurance of monitoring and integrity monitoring. The monitoring procedures include atmospheric monitoring, geochemical monitoring and geophysical monitoring, including seismic surveys. During the injection period, CO2CRC is also monitoring the environment for any changing levels of carbon dioxide in the soil, in the groundwater, in the air and deep underground. The monitoring program started two years before injection period and will proceed for at least two years after the injection process is finished (CO2CRC website, 2009).

IV. DInSAR analysis

In this study, SAR data acquired by the Japanese ALOS satellite were used to monitor the Otway CCS sequestration site. The ALOS satellite was launched on 24 January 2006. It is an L-band SAR system and can be operated in three observation modes: Fine; Polarmetric; and ScanSAR. One Polarmetric image mode used in this study was provided by JAXA (Japan Aerospace Exploration Agency). Specifically, the Polarmetric mode has a resolution of around 20m, an incidence angle of 8-30 degrees at mid-swath, and a total swath width of 70km. Moreover, ALOS is a sun-synchronous polar orbiting satellite which has orbit cycle of 46 days. The available SAR image products of Otway were limited for analysis due to orbit cycle. In addition, the distance between the two satellites (or orbits) in the projection perpendicular to the slant range is called the perpendicular baseline (meter). Besides, temporal baseline refers to the different date between master and slave image (days). The images details are as follows:

Table 1: ALOS PALSAR Data used (a.)

	Satellite	Orbit Cycle (days)	Track	Date yy_mm_dd	Btemp (days)	Bperp (meters)
Master	ALOS-PALSAR	46	387	2007_07_12	598	1463.63
Slave	ALOS-PALSAR	46	387	2009_03_01	598	1463.63

Table 2: ALOS PALSAR Data used (b.)

	Satellite	Orbit Cycle (days)	Track	Date yy_mm_dd	Btemp (days)	Bperp (meters)
Master	ALOS-PALSAR	46	387	2008_08_29	138	-599.402
Slave	ALOS-PALSAR	46	387	2009_01_14	138	-599.402

The targeting 2 pairs were chosen from ALOS PALSAR radar images in order to maximize the amount of CO_2 injection and to keep the balance of both temporal and perpendicular baselines. L-band radar has much less sensitivity to small backscattering changes compared to C-band ERS-1/ERS-2 radar, as is the nature of the longer wavelength in L-band. The two master images used in this study were acquired on 12 July 2007 and 29 August 2008 and the slave images were acquired on 1 March 2009 and 14 January 2009.

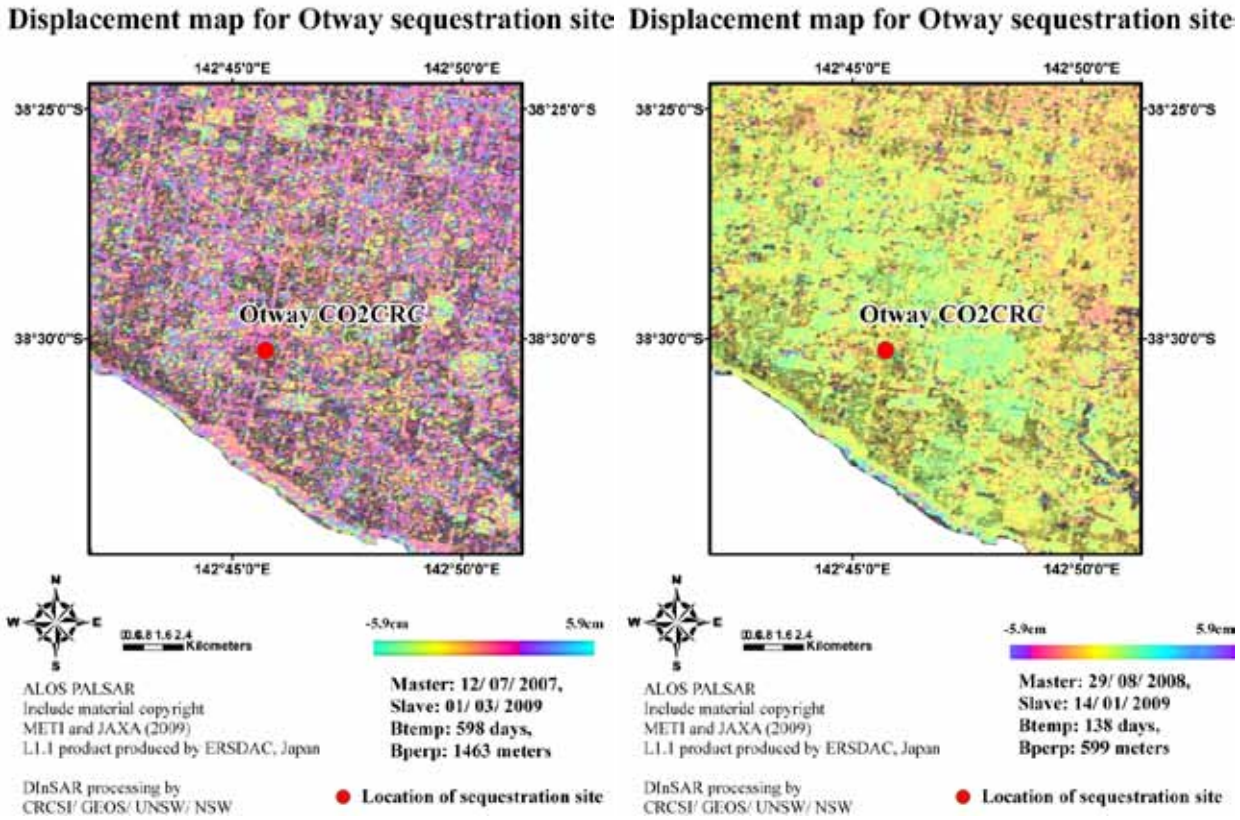


Figure 7: DInSAR result- Otway, Australia

A differential interferograms can be considered to be a contour map of ground displacement along radar Line-of-Sight (LOS) direction. In the case of L-band ALOS PALSAR, each interferometric fringe represents 11.8 cm ground target movement. Figure 7 (a) and (b) are the two differential interferograms generated using SAR images which list in table 1 and table 2, respectively. The two patterns are very similar and there was no ground displacement in the region. The DInSAR processing and interpretation of SAR data show that the atmospheric condition and reflection due to the growth of vegetation could be the main possibility cause which can lead to decorrelation. In this sequestration site, the absent of DInSAR fringes are probably due to atmospheric heterogeneities rather than the decorrelation of changing in vegetation. However, the biases due to the relative change of atmospheric refraction will generally not fail the InSAR processing but seriously mislead the interpretation of the InSAR results. In order to reliably derive ground deformation from the DInSAR results it is crucial to correct for these atmospheric condition first. Besides, some areas had decorrelation due to surface changes and different look angles resulting in large perpendicular baseline and, therefore, cannot be used for measurement of surface deformation.

In addition, it is also important to mention that the CO_2 injection activity will be around 800m to 2 km deep under the ground. And it will be difficult to track subsurface movement of CO_2 which direction is unknown as it is mobility (Harris, 2003) by using conventional monitoring tool such

as filed survey. Therefore, DInSAR is a useful technique which can provide the solution for monitoring the ground surface displacement over a wide area not only around injection well. As the coverage of SAR image will be varies from 50km x 50km to 100km x 100km.

The overall result shows that the surface deformation pattern related with CO_2 injection at Otway, Australia was no potential uplift to be observed in the large area due to several reasons. Firstly, it shows that there is careful site selection in geological characteristics such as the stability of the underground storage reservoir. Furthermore, there are some features which could easily recognize in the image such as road that could also represent stability this area. Secondly, the amount of CO_2 injection may not be massive enough to effect storage formation and appear as ground elevation change. Due to the Otway CO_2 storage site being only a demonstration project and the injection rate is about 100,000 ton through a depleted gas field at a depth of 2 kilometers for two year period (Cook, 2000). Besides, it is possible that the deformation would be hidden by noise. However, if a large ground displacement did occur during both acquisitions (left and right images), as is the case for CO_2 injected, and then it could be detected by DInSAR. Thirdly, if there are greater numbers of interferometry pairs, the deformation time series could be produced which could be part of the supplementary data to build up the underground geological sequestration model.

V. Conclusion and future research

This paper presents the use of radar remote sensing technology which could be one of the monitoring tools for Geologic Carbon Sequestration (GCS) projects. It also concludes that, the amount of injection and the geological characteristic in sequestration storage areas is a big challenge to the successful application of the DInSAR technique. The limited number of SAR images is one of the reasons for the imperfection of the experiment's results.

Further research may consider using a series of DInSAR to improve the results in order to correct the atmospheric condition problem. The ability of this technique also allows the ground deformation to be detected up to millimetres scale. In addition, the Otway sequestration site has a small injection rate, therefore further research may attempt to use satellite data with shorter wavelengths such as X and C band which could detect smaller particles. Besides, a shorter wavelength is needed in order to achieve better sensitivity to ground deformation for radar prospective. It is also expected that the results will be improved by combining with the other monitoring data such as GPS, artificial reflectors and the meteorological satellite images.

Acknowledgements

This study is funded by the Australian Department of Resources, Energy and Tourism (RET) through the Australia-China Joint Coordination Group on Clean Coal Technology Research & Development Grants scheme under the project "Integrated radar and optical satellite remote sensing for safeguarding carbon capture and storage".

The authors also wish to thank the Earth Remote Sensing Data Analysis Centre (ERSDAC) for providing ALOS PALSAR data. METI and JAXA retain the ownership of the ALOS PALSAR original data. The PALSAR Level-1.1 products were produced and provided to the CRC-SI/UNSW by ERSDAC, Japan. Nevertheless, this research was strongly supported by the Geodesy and Earth Observing Systems Group (GEOS) of the School of Surveying and Spatial Information Systems in the University of New South Wales (UNSW).

References

1. Canada global warming 2009, Global Worming, viewed 19 May 2009, < http://www.canadaglobalwarming.com/>.

2. CO2CRC (2004), *Carbon Dioxide Capture & Storage: Research Development & Demonstration in Australia: A Technology Roadmap*, CO2CRC, Canberra.

3. CO2CRC (2009, 2010), CO2CRC Otway Project, viewed 25 June 2009 and 13 January 2011, <http://www.co2crc.com.au/imagelibrary3/otway.php>.

4. Cook P J, Rigg A, Bradshaw J 2000, putting it back where it came from: Is geological disposal of carbon dioxide an option for Australia? APPEA *Journal,* 40 (1), pp. 654-666.

5. Ferretti A, Monti-Guarnieri A, Prati C, Rocca F 2007, InSAR principles: Guidelines for SAR Interferometry Processing and Interpretation (TM-19, February), ESA publications, Netherlands.

6. Ge, L., H-C Chang and C Rizos, 2007. Mine Subsidence Monitoring Using Multi-source Satellite SAR Images, Photogrammetric Engineering and Remote Sensing (PE & RS), 73(3), 259-266.

7. Ge, L., X. Li, C. Rizos, and M. Omura, 2004. GPS and GIS Assisted radar interferometry. Photogrammetric Engineering and Remote Sensing (PE & RS), 70(10), 1173-1178.

8. Harris J M, 2003, Geophysical monitoring of geologic sequestration, *GCEP Technical Report 2003*, pp. 1-5.

9. Harris J M., Quan Y, Wynn D, Ajo-Franklin J, Wu C, Urban J, Akintunde O M, Xu C, *Subsurface Monitoring of Geological CO 2 Storage*, viewed 7 May 2009, < http://gcep.stanford.edu/pdfs/i6W09tDtK-48PTmzHazOuw/3.3.2.Harris_07.pdf>.

10. Mathieson A, Wright I, Roberts D, Ringrose P 2009, 'Satellite Imaging to monitor CO 2 Movement at Krechba, Algeria', *Journal of Energy Procedia*, pp. 2201-2209.

11. International Energy Agency (IEA) 2004, CCS Process Chain (Adapted from Energy Technology Analysis*, prospects from CO_2 capture and storage,* viewed 18 June 2009, <http://www.iea.org/textbase/npsum/ccsSUM.pdf>.

12. International Energy Agency (IEA) 2008, CO 2 capture and storage, viewed 18 June 2009, <http://www.iea.org/Textbase/subjectqueries/cdcs.asp>.

13. IPCC, 2005 - Bert Metz, Ogunlade Davidson, Heleen de Coninck, Manuela Loos and Leo Meyer (Eds.) Cambridge University Press, UK. pp431.

14. Onuma and Ohkawa, 2008 T. Onuma and S. Ohkawa, Detection of surface deformation related with CO_2 injection by DInSAR at In Salah, Algeria, *Proceedings of the 9th International Conference on Greenhouse Gas Control Technologies* GHGT-9, Washington DC, November 16–20 (2008).

15. The National Energy Technology Laboratory (NETL) 2008, Carbon Capture and Storage, viewed 10 June 2009, < http://www.netl.doe.gov/technologies/carbon_seq/refshelf/atlasII/>.

16. Vasco, D W 2002, Geodetic Imaging: Reservoir imaging using satellite interferometry, *Geophysics Journal Int*., 149, pp. 555-571.

17. Vasco, D W, Ferretti, A, Novali, F 2008, *Reservoir monitoring and characterization using satellite geodetic data: Interferometric Synthetic Aperture Radar observations from the Krechba field, Algeria*, paper LBNL'308E, Lawrence Berkeley National Laboratory, University of California, viewed 7 May 2009, <http://repositories.cdlib.org/lbnl/LBNL-308E>.

Design Optimization of a Mars Rover's Rocker-Bogie Mechanism using Genetic Algorithms

Franziska Ullrich, Ali Haydar Göktoğan, Salah Sukkarieh

Australian Centre for Field Robotics (ACFR), J04, University of Sydney, Sydney NSW 2006, Australia

Summary: In this work we present our findings in the optimization of a particular suspension system known as the "rocker-bogie" for our Mars Rover. This type of mechanism has been used on most of the rovers on Mars and has proved to be a simple and elegant design. This work defines the most important performance metrics for a planetary rover. A Genetic Algorithm is implemented and used to optimize the geometry and kinematics of the rover's wheel suspension system subject to the defined performance metrics. A Genetic Algorithm is chosen for optimization due to its simplicity of implementation and fast convergence to an optimum within a complex and unknown search space.
This work shows the effectiveness of the optimization of a rocker-bogie suspension system using a Genetic Algorithm. It also reveals that the resulting system meets all constraints and that significantly reduces the error of individual performance metrics and the overall system. It is shown that the overall fitness of the rover suspension system can be increased by an average of 28% after 100 iterations compared to an initial guess. All performance metrics defined are improved significantly throughout the optimization.
The method can be applied to different types of rovers in order to optimize the wheel suspension mechanism's geometry.

Keywords: Mars Rover, Rocker-Bogie Mechanism, Design Optimization, Genetic Algorithm

Introduction and Background

The type of locomotion used by a mobile robot is crucial for the robot to perform its task and reach its goal in a given environment. This work focuses on the optimization of the design of a planetary rover's wheel suspension system subject to optimizing well defined mobility metrics.
As robots evolve from industrial fixed base robots to autonomous mobile platforms, the concept of locomotion in robotics becomes much more important. Similar to nature, also robot locomotion must be adapted to the given terrain or task. The optimal type of locomotion must be applied in a challenging environment.
The scope of this work is to design an efficient Mars Rover suspension system and to develop and implement a Genetic Algorithm methodology which optimizes the design of the locomotion system and can be applied to diverse mechanisms or other problems.
As the Mars Rover is a mobile robot, the wheel suspension system of the rover is most crucial. It allows for movement, mobility and stability of the robot while it is travelling through a Mars environment. The rover must be able to traverse over obstacles of at least half its wheel diameter and keep its stability on slopes or other rough or hazardous terrain. Hence, the

suspension system is to be optimized using a Genetic Algorithm.
A planetary rover is a mobile robot which is especially designed to move on a planet surface. Early rovers were teleoperated like the Lunokhod I [1] while recent ones are fully autonomous, such as FIDO [2]. The rover has to be very robust and reliable, as it has to withstand dust, strong winds, corrosion and large temperature changes. Most rovers are powered by batteries which are recharged by solar panels during the day. Therefore, the rover must position itself in such a way that the solar energy received is maximized. The rover's locomotion system is crucial to enable it to reach a goal, conduct experiments, gather data and to position itself. Generally three main types of rover locomotion exist, namely wheeled, legged and caterpillar locomotion [3], [4]. The key difference between the miscellaneous designs of planetary robots lies in the type of locomotion system. Even though many legged and hybrid robots have been presented in literature, most researchers still focus on wheeled locomotion for rovers. The favoured design for a wheeled planetary rover's suspension system is the rocker-bogie mechanism. Numerous variations of the this mechanism have been presented in literature. The FIDO rover [2] and the Sojourner [5] have six independently steered and driven wheels suspended from a rocker-bogie mechanism. The Rocky7 Rover [6] has a similar suspension system, but only the front wheels are steered. The Nanorover [7] and the Nomad [8] have four steered wheels suspended from two bogies, a variation of the rocker-bogie mechanism. The CRAB Rover [9] uses two parallel bogie mechanisms on each side to overcome obstacles.
Lamon et al. [10] optimize a simplified quasi-static model of the six-wheeled rover Shrimp. The motor torques are optimized subject to minimizing the slip of the wheels, hence the odometric error and power consumption are minimized. The optimization seeks for an optimum in the constrained solution space given an initial solution. Li et al. [11] derive a mathematical model to optimize rover suspension parameters which define the geometry of the rocker-bogie. The objective is to minimize the energy consumption, the vertical displacement of the rover's centre of mass and its pitch angle. The authors make use of a sequential quadratic programming algorithm. Iagnemma and Dubowsky [12] introduce an algorithm which optimizes individual wheel ground contact forces. The objective is to optimize traction and to minimize power consumption.

Rocker-Bogie Modelling

The Rocker-Bogie Mechanism

The rocker-bogie suspension system was first used for the Mars Rover Sojourner and is currently NASA's favoured design for rover wheel suspension. The intelligently designed wheel suspension allows the vehicle to traverse over very uneven or rough terrain and even climb over obstacles. The rocker-bogie allows the chassis of the rover to average its pitch over all wheel deflections, while still maintaining load equalization on all wheels and avoiding a low oscillation frequency.
The suspension has symmetric structure on both sides of the rover chassis and supports a total of six wheels, three on each side. The mechanism is constructed from two rigid parts. The link connected with a rotary joint to the chassis is called the *rocker*. The *bogie* is connected to the rocker via another rotary joint, as shown in the kinematics diagram in Fig. 1. The rockers are connected to each other via a differential mechanism, which allows the rover body to minimize its pitch.
The rocker-bogie suspension mechanism has two main advantages. The total load that is put

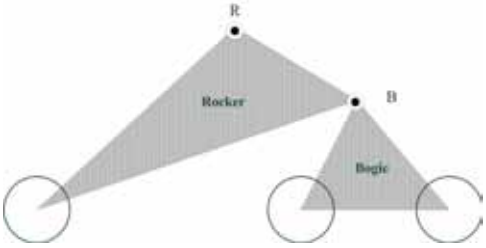


Fig. 1: Kinematics Diagram of a Rocker-Bogie Mechanism

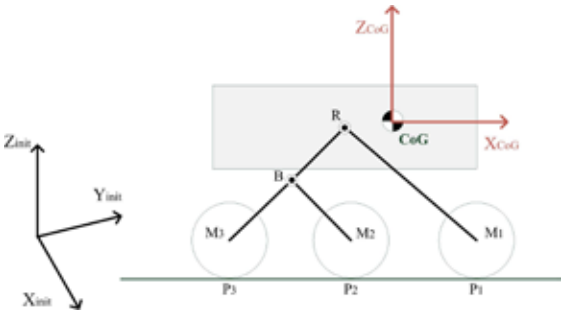


Fig. 2: Points M_i, P_i, B, R and Reference Frames and their Locations on the Rocker-Bogie

on the rover is distributed evenly over all wheels. This ensures equal working conditions for all wheels and prevents from excessive sinkage of wheels in soft terrain [13]. The second advantage is that all wheels stay in touch with the ground while climbing obstacles. This ensures stability and prevents the rover from tilting on a slope. The rocker-bogie is a passive suspension system and has no axles or springs, all wheels are kept in contact with the ground without any actuators, decreasing power consumption of the rover.
Each of the six wheels is independently driven and steered. This also allows the rover to turn in place or move sidewards. The system therefore is redundant and can move forwards, backwards and sidewards.

Simplified System Model

The mathematical model basically takes into account the entire geometry of the rocker-bogie, in terms of link lengths and coordinates of connection points. Instead of concentrating on the exact geometry, it is important to make use of all connection points between links. Therefore, a L-shaped link can be constructed in another way in the realistic model as long the end points are kept. The exact shape of the links is then to be designed according to the general skeleton, which is subject of the latter geometry optimization.
In the optimization process the coordinates of the wheel midpoints (M_i), the coordinates of the wheel terrain contact points (P_i) and the coordinates of both rocker (R) and bogie (B) connection points are used, where $i = \{1 \dots 6\}$ is the number of wheels. All these coordinates are computed with reference to a global frame ($X_{init}, Y_{init}, Z_{init}$) and the body-fixed frame ($X_{CoG}, Y_{CoG}, Z_{CoG}$), located at the centre of gravity (CoG). The forward kinematics of the rover are derived and the coordinates of points R, B, P_i and M_i are found, shown in Fig. 2.

When deriving the forward kinematics of the rocker-bogie suspension system, it is assumed that the terrain and its slope are known. Also the gradient of the ground is known at every point of the terrain. This is not a realistic assumption, as in the real world it is a very challenging task to find the terrain's shape. Furthermore, a point contact between wheel and terrain is assumed. This assumption is true for a rigid wheel and a relatively firm ground.
Another assumption is that the rover does not experience sideslip. Furthermore, it is assumed there are no moments acting at the wheel-terrain interface. In this work, all links are assumed to be rigid bodies. The maximum travelling speed for the rover is assumed to be $0.25\ \frac{m}{s}$,

the maximum acceleration is assumed to be $1\ \frac{m}{s^2}$. As maximum velocity and acceleration are relatively small, the rover can be modeled as a quasi-static system and all dynamic effects are neglected [10]–[12].

Performance Evaluation and Metrics

Performance evaluation of a system is a necessary process, which provides many information about the system and helps to improve future systems. Prior to an optimization of the rocker-bogie suspension system, it is important to define the performance metrics. The performance metrics form the basis for the objective function, which is to be maximized during the optimization process.

Wheel-Terrain Contact Forces

The analysis of the rover's suspension system's design demands for the knowledge of the wheel-terrain contact interaction. It is assumed, that each rigid wheel makes contact with the known and relatively firm ground at a single point, P_i. The vectors from the contact points P_i to the rover's centre of gravity are given by $\mathbf{p_i} = [p_i^x, p_i^y, p_i^z]^T$ as illustrated in Fig. 3. The contact force vector $\mathbf{f_i} = [T_i, N_i, S_i]^T$ where T_i is the traction force at wheel-terrain contact point P_i, N_i is the normal force and S_i is the force acting sidewards. The vectors $\mathbf{p_i}$ and $\mathbf{f_i}$ are expressed in the rover's bodyframe.
The 6×1 vector $\mathbf{f_s}$ is expressed in the rover's bodyframe and is located at the rover's centre of gravity. It represents all forces and moments acting on the rover body, such as gravitational forces, inertial forces, manipulation and driving forces. The decomposition of $\mathbf{f_s}$ is shown in Fig. 4, where α_i is the angle of incline of the ground at the wheel-terrain contact point P_i. As suggested by Iagnemma and Dubowsky [12] a set of quasi-static force balance equations are written as

$$\begin{bmatrix} \mathbf{I} & \cdots & \mathbf{I} \\ 0 & p_1^z \quad p_1^y & \cdots & 0 \quad p_6^z \quad p_6^y \\ -p_1^z & 0 \quad -p_1^x & \cdots & -p_6^z \quad 0 \quad -p_6^x \\ -p_1^y & p_1^x \quad 0 & \cdots & -p_6^y \quad p_6^x \quad 0 \end{bmatrix} \begin{bmatrix} \mathbf{f_1} \\ \vdots \\ \mathbf{f_6} \end{bmatrix} = \mathbf{f_s} \tag{1}$$

in order to find the wheel-ground interaction forces. In (1) $\mathbf{I}$ is the 3×3 identity matrix and all dynamic effects are neglected.
The matrix equation (1) can be rewritten as $\mathbf{Gx} = \mathbf{f_s}$, where $\mathbf{G}$ is a function of the rover's geometry, the wheel-terrain contact location and the wheel-terrain contact angles α_i. The vector $\mathbf{f_s}$ is the input vector of all forces acting on the rover. The equation is solved for $\mathbf{x} = [T_1, N_1, S_1 \ldots T_n, N_n, S_n]^T$, $n = 6$, using the Moore-Penrose pseudo inverse, defined as $\mathbf{x}^+ = (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\mathbf{f_s}$. For this calculation matrix $\mathbf{G}$ is assumed to have full rank. This approach is very inexpensive, however, it might not satisfy the system's physical constraints.

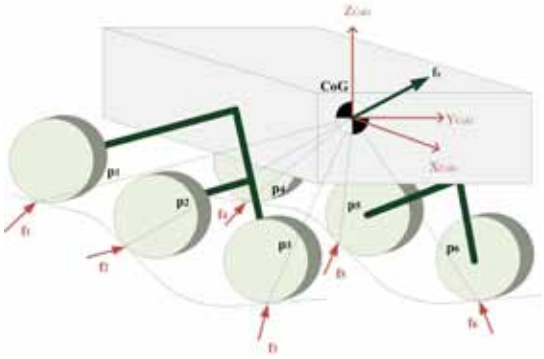


Fig. 3: Rover with six Wheels on Rough Terrain

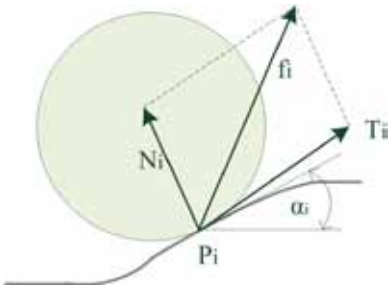


Fig. 4: Derivation of Forces at Wheel-Terrain Contact Point

Power Consumption

As a planetary rover can not recharge via a power plug or change batteries during its mission, it is very important that power consumption of the rover is low. Power consumption is mainly the energy being used by the driving and steering motors. As suggested by [12], it is given by $P = \frac{R_M g_M^2}{K_t^2} \sum_{i=1}^{n} \Gamma_i^2$, where R_M is the motor resistance, g_M is the motor gear ratio, K_t is the motor torque constant and Γ_i is the torque applied to the *i*th motor. As the applied torque is the product of traction force T_i at wheel *i* and the same wheel's radius r, power consumption can be written as $P = \frac{R_M g_M^2 r^2}{K_t^2} \sum_{i=1}^{n} T_i^2$. It is obvious that power consumption is proportional to the sum of all squared traction forces T_i. In order to minimize the power consumption of the rover during travel, the sum of the squared traction forces is to be minimized.

Traction and Slip

The rover must maintain good wheel traction in challenging rough terrains. If traction is too high, the vehicle consumes a lot of power in order to overcome the force and move. If traction is too low, the rover is not able to climb over obstacles or inclined surfaces. Slip occurs when the traction force at a wheel-terrain contact point is larger than the product of the normal force at the same wheel and the friction coefficient. Hence, no slip occurs if the condition $T_i \leq \mu N_i$ is satisfied. In reality it is very challenging to determine the precise friction coefficient μ for the interaction of two surfaces. As an approximation of the friction coefficient, a virtual friction coefficient μ^* is computed, according to Lamon et al. [10], as the ratio between traction and normal force at one wheel-terrain contact point.

$$\mu^* = \max_i \{\frac{T_i}{N_i}\} \tag{2}$$

The parameter μ^* is similar to a friction coefficient. If the constraint $\mu^* \leq \mu$ is satisfied, no slip occurs. Hence, μ^* is targeted to be minimized. According to [10] and [14] the optimum solution is found if μ^* is equal for all n wheels.

Load Equalization

Load equalization over all n wheels ensures equal work conditions for all wheels. The following Fig. 5 shows the two dimensional mechanics model of the rocker-bogie suspension. Weights G, normal forces N and traction forces T are shown as well as geometric dimensions. The weight G_b is the total weight of the rover body while G_i denotes the respective weight from the rocker-bogie mechanism acting on wheel i. A quasi-static force balance is conducted, as suggested by Li et al. [11], and the conditions $x_B = \frac{3}{2}x_C$ and $x_B = \frac{1}{2}(x_1 + x_2)$ are derived for load equalization. These geometrical constraints for the suspension structure are used to maximize load equalization.

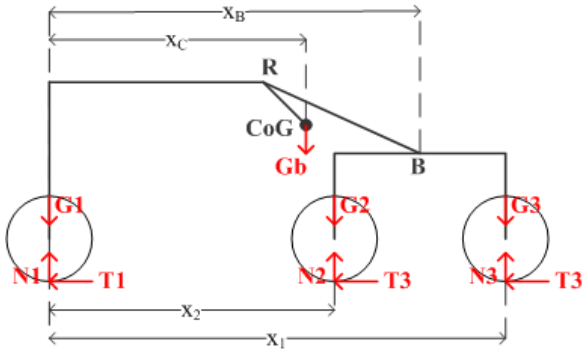


Fig. 5: Load Equalization over all n Wheels

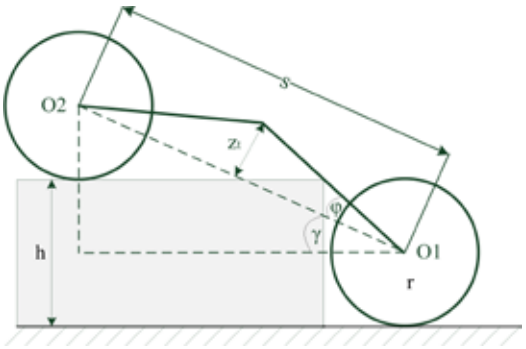


Fig. 6: Geometric Trafficability of the Bogie

Geometric Trafficability

Geometric trafficability ensures that the suspension structure does not interfere with the ground. The analysis of geometric trafficability is only done for the bogie, as the swinging scope for the bogie is much bigger than that of the rocker. The analysis of geometric trafficability is done according to [11]. From geometry the maximum vertical distance z_t between the centre of the wheels to the edge of the obstacle is determined according to (3), where r is the wheel radius and h and s are geometric properties, seen in Fig. 6.

$$z_t = \frac{s}{\sqrt{\left(\frac{r\sqrt{s^2-h^2}+(h-r)h}{(h-r)\sqrt{s^2-h^s}-hr}\right)^2 + 1}} \tag{3}$$

$$z_t \leq c \tag{4}$$

From geometry the maximum vertical distance z_t between the centre of the wheels to the edge of the obstacle is determined. Geometric trafficability is given if the parameter z_t is larger or equal to the rover's ground clearance.

Lateral and Longitudinal Stability

The rover is said to be stable when it is in a quasi-static state in which it does not tilt over. The simplest approach to find the static stability is using the geometric model, which

is commonly referred to as stability margin. As the asymmetric suspension system of the passively articulated rover has a great influence on the vehicle's effective stability, a more advanced approach is using a static model. The geometric a) and the statical b) model for the rocker-bogie suspension system are illustrated in Fig. 8.

The lateral stability of the rover ensures that the rover does not tip sideways. As the rover has

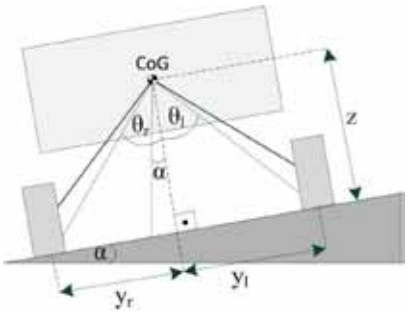


Fig. 7: Lateral Stability of the Rover

Fig. 8: Stability Models for Evaluation

two symmetric sides, the geometric model is used to find the lateral stability of the vehicle, as suggested by Thüer [14]. Lateral stability is computed by finding the minimum allowed angle on the slope before the rover tips over. Lateral stability is ensured if this angle is smaller than the maximum angle of incline α on the slope at the wheel-terrain contact points. The angles θ_l and θ_r are obtained geometrically. The overall stability angle θ_{stab} can be computed by

$$\theta_{stab} = \min\left(\theta_{\mathrm{r}}, \theta_{\mathrm{l}}\right) \tag{5}$$

Lateral stability of the rover is ensured if the overall stability angle $\theta_{stab} \geq \alpha$.

The computation of the longitudinal stability of the rover makes use of a statical model as it is not symmetric in longitudinal direction. Using a statical model, the mechanical properties of the suspension system are taken into account. According to [14], longitudinal stability of the vehicle is given when all wheels have ground contact and the condition $N_i > 0$ is satisfied, where N_i is the normal force at wheel i. It should be noted that even though this condition is compulsory for the statical model to work, a physical rover does not necessarily tip if a wheel looses contact to the ground. However, it is less steerable.

Sinkage

A rover traveling over terrain is subject to sinkage, depending on the properties of the soil as well as loading, geometry and shape of the wheels. Apostolopoulos [15] suggests an experimentally derived formula for the maximum sinkage experienced by a rigid wheel in weak soil. The soil is defined by soil parameters which have been tabulated by Wong [16] for various soils. An optimization of the vehicle's geometry seeks to reduce the maximum sinkage of the wheels on soil.

Pitch Variation

As the rover is powered by solar energy, its solar panels should be perpendicular to the sun rays at all times. Therefore, very low variation in the rover's pitch angle is desirable.

A differential mechanism connects the two rocker-bogies. As the rover traverses over rough terrain its pitch is the average of the two rocker deflections. By choosing an appropriate gear ratio the overall pitch of the rover is optimized.

Design Optimization of the Rocker-Bogie

This work uses a Genetic Algorithm in order to find an optimal solution for the rocker-bogie suspension. The Genetic Algorithm shows some advantages over other optimization techniques. It optimizes with discrete or continuous variables and does not require derivative information. Also, it optimizes a large number of variables with extremely complex search spaces and is therefore a global search algorithm.
The Genetic Algorithm is started with an initial population, which consists of a number of chromosomes, each representing a possible solution to the optimization problem. The aim of the optimization is to minimize the cost and hence maximizing the *fitness*. Each chromosome is associated with a certain fitness, given by the fitness criterion or objective function. In order for the population to evolve, in each generation *selection* of the parent chromosomes takes place. These parents then pair up to produce new offspring that replaces the old population. Different methods for selecting parents are random selection, roulette wheel selection, ranked selection and tournament selection. Discarding of all chromosomes in each new generation might not be very efficient, when searching for the best solution. Therefore, the principle of *elitism* is introduced. This means that the fittest chromosome of each population is kept. Hence, the best solution of the last population of chromosomes is copied to the new population and is ensured to survive. The process of *mating* is the creation of new offspring from the parents, selected before. Every two parents produce one new offspring by the means of *crossover*. The first parent passes on some properties while the second parent passes on others. The new offspring consequently has parts of both parents. It is inserted into the new population and might be chosen to be a parent chromosome in the next generation, according to its fitness. *Mutation* is a way to explore the search space. While mating is limited by those chromosomes in the population, mutation can introduce new traits in the population. These new attributes may increase a population's fitness or may decrease it. Increasing the number of mutations distracts the algorithm from converging to one popular solution, and hence makes the Genetic Algorithm a global search algorithm. The generational process is terminated when the termination criterion is reached.

During the rocker-bogie optimization the geometry of the simplified model is to be optimized subject to the objective function, shown in equation (6)

$$\begin{aligned} \text{fitness} \;=\; & -w_1 s\mu^* - w_2\varepsilon_{\mu^*} + w_3\,(s-1)\,P + w_4 c_{lat} + w_5 c_{long} \\ & -w_6\,(\varepsilon_1+\varepsilon_2) + w_7 c_{traff} - w_8 z_{rw} - w_9\theta_{rover} \end{aligned} \tag{6}$$

where w_1 to w_9 are the weights, as shown in Table 1. Parameter μ^* is the virtual friction coefficient. By minimizing the error ε_{μ^*} the difference of virtual friction coefficient at each

Table 1: Weights for Performance Metrics during Optimization Series

Weight no.	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9
Weight	2	1	2	1	5	3	2	1	1

Table 2: The Upper (UL) and Lower Limits (LL) of the Optimization Variables

Variable	x_R	y_R	z_R	γ_{RB}	x_{B1}	y_{B1}	x_{B2}	y_{B2}	c	L_{RB}	j
LL $[mm]$	100	100	100	90$[deg]$	100	100	100	100	20	50	1$[-]$
UL $[mm]$	500	300	200	180$[deg]$	200	300	200	300	100	500	5$[-]$

wheel is decreased, leading to similar work conditions for all wheels. Parameter P is proportional to the power consumption. In rough terrain, traction should be maximized, while in benign terrain, power consumption should be minimized. The switching function s ensures this property and is given by

$$s = \begin{cases} 1, & \text{if } \max(|\alpha|) > C \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

where α_i is the terrain's angle of incline at each wheel-terrain contact point. C is an arbitrary threshold, which defines what is considered rough or benign terrain. Furthermore, c_{lat}, c_{long} and c_{traff} are parameters, which are large $(= 1000)$ if lateral stability, longitudinal stability and trafficability are given, respectively, and zero otherwise. Parameters ε_1 and ε_2 are to be minimized, in order to ensure load equalization. The sinkage z_{rw} and the rover pitch θ_{rover} are minimized.
Geometrical parameters of the rocker-bogie model as well as the gear ratio j of the differential mechanism connecting both rocker-bogies are to be optimized and represented in the chromosome, as illustrated in (8) and shown in Fig. 9.

$$\text{chromosome} = [\ x_R\ \ y_R\ \ z_R\ \ \gamma_{RB}\ \ x_{B1}\ \ y_{B1}\ \ x_{B2}\ \ y_{B2}\ \ c\ \ L_{RB}\ \ j\ ] \tag{8}$$

All geometrical parameters are constrained by real numbers representing the physical constraints, these are shown in Table 2.

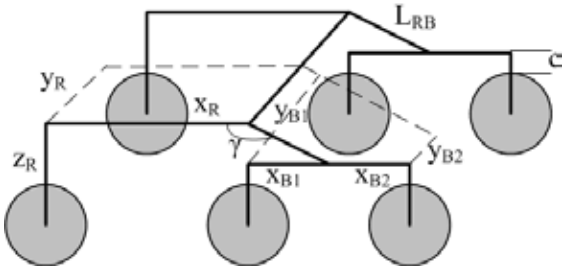


Fig. 9: Geometry Model of Rocker-Bogie showing all Geometrical Parameters

During optimization the ground profile assumed is a hill with specified angle of inclination α, but can be changed by the user.
The Genetic Algorithm evaluates the fitness of each chromosome in the population and therefore quantifies the goodness of each of these specific combinations of link lengths and angles of the rocker-bogie mechanism. It then mates two good solutions in order to find even better combinations of variable values. The fitness is determined by the objective function, which includes all the performance metrics described above. The traction is optimized while the slip and power consumption of the rover are minimized. Furthermore, load equalization is maximized and sinkage of the wheels and the variation of the rover's pitch are minimized. Also, geometric trafficability, lateral and longitudinal stability are to be ensured by minimizing the objective function and hence maximizing the fitness of the best solution, represented by the best chromosome in the final population. All performance metrics are weighted in the objective function. By adjusting the weights, the user can emphasize one metric's importance over another one's. By maximizing the fitness function, the performance of the rover is enhanced. As fitness is a function of the rover suspension system geometry, the optimum for the geometrical parameters can be found.

Results

The principles of elitism and mutation are applied with a mutation rate of 0.2. The selection type chosen for this optimization is tournament selection. The rigid wheels have a diameter of $d_w = 170\ mm$ and a width of $b_w = 75\ mm$. The rover is moving over sand and the maximum obstacle height to be climbed is $h = 170\ mm$, equal to the wheel diameter. The Genetic Algorithm is terminated after 100 generations. The optimization algorithm is run 25 times during the series. Fig. 10 shows the progression of the best fitness within one population, representing 70 chromosomes, for 100 generations. Fig. 11 visualizes the progression of the average fitness within the same population. All performance metrics are weighted as shown in Table 1. Both, Fig. 10 and 11, show an example regression which is similar for all 25 runs. The values found for the optimization parameters are different for every run. No two solutions found are exactly the same. Nonetheless, most values are in a similar range and show a specific trend for the particular dimension.

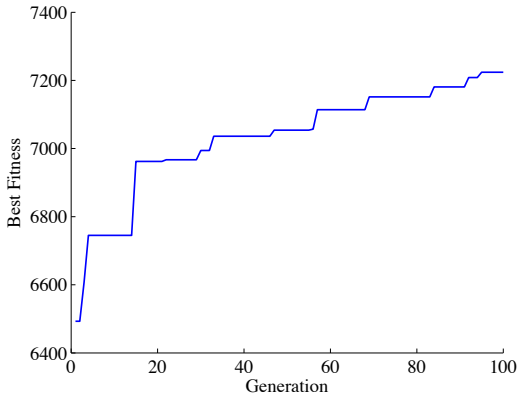


Fig. 10: Progression of the Best Fitness during Optimization over 100 Generations

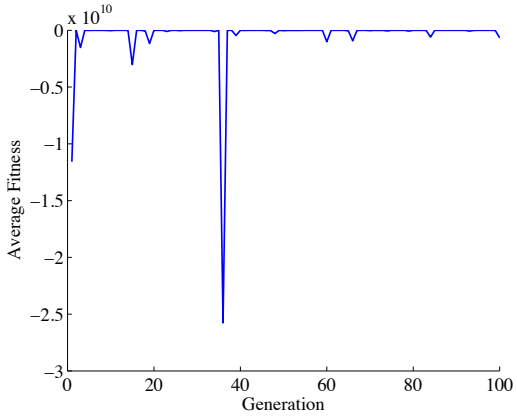


Fig. 11: Progression of the Average Fitness during Optimization over 100 Generations

Analysis and Discussion

In general, it is shown that the Genetic Algorithm significantly increases the fitness of the chromosome over 100 generations and therefore optimizes the geometry of the rover's rocker-bogie suspension system subject to the performance metrics described.
Fig. 10 illustrates the best fitness within the evolving population, which represents a set of 70 different solutions. It is clear, that the fitness is increased stepwise. The shape of the progressing fitness is given due to the principle of elitism. The best fitness is that of the best solution which is kept in the evolving population until a better solution with a higher fitness is found. So the solution with the highest fitness is exploited, as the value for best fitness does not decrease during the optimization. The results show a large increase of the best fitness within the population of possible solutions between generation one and 100. While absolute increase of the fitness is 1527 on average compared to a random guess within the physical constraints, the relative increase is between 7% and as much as 104%, with an average of 28% among the 25 runs. The optimization algorithm improves the fitness significantly over 100 iterations.

Fig. 11 shows the progression of the mean fitness within one population. It is obvious that many areas of the search space are explored as the mean value for the fitnesses oscillates over a wide range. The principle of mutation makes the optimization a global search within a large search space. As point mutations are inserted into a new population, other parts of the search space are explored.
The optimal solution found during each run is different to the others, but the optimized values found for each parameter, represented by the chromosome, are in a similar range and show a trend for each optimization parameter. This suggests that multiple designs for the rocker-bogie show similar values for the fitness and are equally good in terms of optimizing the performance metrics. Hence, the dimensions of the suspension system, such as link lengths and angles between links, are strongly interdependent. In general, the rocker-bogie mechanism is found to be optimal for certain structural details. The total width of the rover is found to be optimal if it is large, while total height should be small. Both of these details can be explained by the desire for lateral and longitudinal stability of the rover. The connection point between rocker and rover body is not located in the centre but to the front of the rover. The angle between rocker and bogie is large, while the connecting link is rather short. This ensures a long vehicle but less weight than for a long connecting link. The middle and rear wheel are not in line with the front wheel for an optimal solution but are located further from the rovers centre. This detail also increases stability, especially during traversing over an obstacle. Furthermore, the ground clearance is kept as small as possible while still ensuring geometric trafficability of the rover. After optimization of the total fitness of a solution, each performance metric was computed separately using the optimal values for the rocker-bogie geometry found by the Genetic Algorithm. Validating each individual performance metric shows that the quality of every metric was increased during optimization by a large percentage. Furthermore, lateral stability, longitudinal stability and geometric trafficability are given in 100% of the cases. The results show that not only the total quality of the rocker-bogie mechanism was increased during the optimization, but also each performance metric could be optimized using a Genetic Algorithm.

Conclusion

This work describes the modelling of a quasi-static rocker-bogie suspension system for a Mars Rover, followed by the identification of several performance metrics which are to be optimized if the rover is to execute its tasks and reach its goals in the best possible way. Subsequently, an optimization method is implemented. The applied method is a Genetic Algorithm which is an evolutionary computing method.
Several performance metrics have been introduced and described which give a thorough basis for the evaluation of the rover's rocker-bogie suspension system. Furthermore, a Genetic Algorithm is described and applied to the optimization problem. The results of the optimization and the analysis of the same shows that the rover's suspension mechanism's geometry is optimized. After optimization it shows a better performance and its abilities to traverse are improved. Overall performance is increased by an average of 28% as well as the individual performance metrics.

Acknowledgements

I thank Stefan Würgler and Taufik Yunahar for their helpful comments and suggestions.

References

[1] Cherkasov, II and Shvarev, VV, "Soviet investigations of the mechanics of lunar soils", *Soil Mechanics and Foundation Engineering*, Vol.10, No. 4, 1973, pp. 252-256.

[2] Schenker, PS and others, "FIDO: A field integrated design & operations rover for Mars surface exploration", *i-SAIRAS, Montreal, Canada*, 2001.

[3] Siegwart, R. and Lamon, P. and Estier, T. and Lauria, M. and Piguet, R., "Innovative design for wheeled locomotion in rough terrain", *Robotics and Autonomous systems*, Vol. 40, No. 2-3, 2002, pp. 151-162.

[4] Estier, T. and Crausaz, Y. and Merminod, B. and Lauria, M. and Piguet, R. and Siegwart, R., "An innovative space rover with extended climbing abilities", *Proceedings of Space and Robotics*, 2000, pp. 201-206.

[5] Mars Pathfinder Rover by NASA, http://nssdc.gsfc.nasa.gov/nmc/spacecraftDisplay.do?id=MESURPR, 2009.

[6] Rocky 7: A next generation mars rover prototype, "Rocky 7: A next generation mars rover prototype", *Journal of Advanced Robotics*, 1997.

[7] Mars Pathfinder Rover by National Aeronautics and Space Administration (NASA), http://www-robotics.jpl.nasa.gov/tasks/showTask.cfm?FuseAction=ShowTask&TaskID=109&tdaID=888881, 2009.

[8] Wettergreen, D. and Bapna, D. and Maimone, M. and Thomas, G., "Developing nomad for robotic exploration of the atacama desert", *Robotics and Autonomous Systems*, Vol. 26, No. 2-3, 1999, pp. 127-148.

[9] Thüer, T. and Lamon, P. and Krebs, A. and Siegwart, R., "CRAB-Exploration rover with advanced obstacle negotiation capabilities", *Proceedings of the 9th ESA Workshop on Advanced Space Technologies for Robotics and Automation*, 2006.

[10] Lamon, P. and Krebs, A. and Lauria, M. and Shooter, S. and Siegwart, R., "Wheel torque control for a rough terrain rover", *IEEE International Conference on Robotics and Automation*, Vol. 5, 2004, pp. 4682-4687.

[11] Li, S. and Gao, H. and Deng, Z., "Mobility performance evaluation of lunar rover and optimization of rocker-bogie suspension parameters", *Systems and Control in Aerospace and Astronautics, 2008. ISSCAA 2008.*, 2008, pp. 1-6.

[12] Iagnemma, K. and Dubowsky, S., "Traction control of wheeled robotic vehicles in rough terrain with application to planetary rovers", *The International Journal of Robotics Research*, Vol. 23, No. 10-11, 2004, pp. 1029.

[13] Harrington, B.D. and Voorhees, C., "The challenges of designing the rocker-bogie suspension for the Mars Exploration Rover", *37th Aerospace Mechanisms Symposium*, 2004.

[14] Thüer, T., "Mobility Evaluation of Wheeled all-terrain Robots - Metrics and Application", *PhD Thesis, Swiss Federal Institute of Technology Zurich (ETH Zurich)*, 2009.

[15] Apostolopoulos, D., "Analytical configuration of wheeled robotic locomotion", *Unpublished Ph. D. thesis, The Robotics Institute, Carnegie Melon University, Pittsburgh, PA*, 2001.

[16] Wong, J.Y., *Theory of ground vehicles*, 2001.

Path Planning for a Planetary Rover

Stefan Würgler *, Salah Sukkarieh *

* *Australian Centre for Field Robotics, University of Sydney, NSW, Australia, 2006*

Summary: Two state of the art path planning algorithms, Field D* and State Lattice, are analysed regarding their suitability for all-wheeled steered rovers. A new rover adapted path planning approach (raD*) is then presented. The generated paths of the three algorithms are compared regarding computational efficiency, feasibility, smoothness, and required acceleration and steering action. The proposed raD* path planner creates paths that are smoother trackable than the ones extracted from the Field D* cost map. Compared to State Lattice, raD* generates paths with similar smoothness, but requires an order of magnitude less computational power. As tracking the paths also involves less steering action, we state that raD* is superior to Field D* and State Lattice for all-wheel steered rovers.

Keywords: Path planning, State Lattice, Field D*, raD*, all-wheel steered rover, differential constraints

Introduction

Planetary rovers are designed to explore the unknown environment of a remote planet. The latency in the communication between the earth and the rover, caused by the large distance, makes a teleoperation of the rover infeasible. Autonomy of the rover increases the range of missions that can be performed without human interaction. Path planning is a crucial part of autonomous operations, in order to navigate the rover to areas of interest while circumnavigating obstacles and avoiding hazardous areas.

Path planning for planetary rovers are usually performed on grid maps to reduce the computational complexity of the problem. Standard search algorithms on grid maps result in suboptimal paths and are limited to a few headings, so that the resulting paths include unnecessary sharp turns. To overcome this problem, more advanced techniques such as Field D* were proposed, which can generate paths with arbitrary headings using interpolation between neighbouring cells [1]. Field D* is currently used on NASA's Mars Exploration Rovers (MERs) [2]. As the rover's differential constraints are not taken into account on the global level, non-feasible paths are only omitted by a local path planner later in the process. A more recent approach using "state lattices" generates paths with a set of path segments which satisfy the vehicle's differential constraints in the first place [3].

While most rovers are capable of turning in place, current all-wheel steered research rovers, such as Rocky 8 and ExoMars, can additionally drive sideways. Consequently, an arbitrary path can be followed by such rovers. The main advantage of the State Lattice approach, that it only generates feasible paths, does therefore no longer apply. This work explores if its nevertheless beneficial to impose artificial mobility constraints in order to generate smooth paths for all-wheel steered rovers using State Lattice.

A new rover adapted path planning algorithm (raD*) is then proposed that creates paths that are smoothly trackable without limiting the rover's motion capabilities. Finally, the different path planning approaches are tested on various scenarios. The resulting paths are compared regarding computational efficiency, feasibility, smoothness, and required acceleration and steering action.

Path Planning Algorithms

Map Representation

Most path planning algorithms in mobile robotics use a two dimensional grid map to represent the environment. Each cell of the grid has a cost value assigned that rates the traversability of the terrain, where low values indicate easy traversable cells and high values hazardous areas. Generally the map is constructed in configuration space, where obstacles are expanded by the rover radius, so that the rover can be treated as a point in later computations.

The grid map builds a graph structure, where each cell is connected to its 4 or 8 adjacent cells. This discretization of space improves the computational efficiency, but it sacrifices optimality and feasibility. Due to graph structure itself, "optimal" graph paths can be suboptimal with respect to the real world. The path is limited to horizontal and vertical segments for 4-connected graph structures, and headings of multiples of $\pi/4$ for 8-connected graphs. This small, discrete set of possible transitions artificially constrains the rover's movements. This can result in paths that are suboptimal in length and involve unnecessary turns. Paths with instantaneous heading changes can be difficult or even infeasible to follow in practice, depending on the rover's motion constraints.

D* & D*Lite

Given the discretized map as a directed graph, any graph search algorithm can be used to find a path from the start node to the goal point. By focusing the search with a heuristic estimate of the distance to the goal, A* finds the optimal path efficiently. In mobile robotics the map is continuously updated, as additional information is gathered while driving around. As the edge costs of the graph change accordingly, the path has to be replanned repeatedly. By reusing information from previous searches, incremental heuristic search algorithms, such as D* and its variants, are able to efficiently recompute the path when changes in the underlying graph occur. Most of the current systems run with a version of D*Lite, which was developed by Koenig and Likhachev [4].

Field D*

Field D* was introduced by Ferguson and Stentz [1] to address the problem caused by the small set of discrete state transitions resulting in unnatural paths limited to a few headings. Field D* is an interpolation-based planning and replanning algorithm, that produces paths with non-discretized headings. Path segments can no longer only connect the midpoints of grid cells, but can intersect the grid cell boundaries at any point. Field D* is currently

implemented on several field robotic systems, amongst others on NASA's Mars Exploration Rovers Opportunity and Spirit [2].

Finding the optimal path without restricting the translations to the grid cell midpoints is intractable, as there is an infinite number of points along the boundary of the grid cell. Assuming the cost of a point on the boundary is the linear combination of the costs of the adjacent corner points, the cost of any boundary point can be approximated. The total cost of a given node is then determined by the minimum of the sum of the travel cost through the cell to a boundary point and the interpolated cost of that boundary point. Thanks to the linear interpolation assumption, this minimization problem has a closed-form solution. Given this interpolation based cost calculation for every node, the cheapest path can be calculated by any current graph search algorithm. Ferguson and Stentz [1] present an implementation of Field D* based on D*Lite.

Extracting paths from cost fields produced with Field D* is not straightforward. For simple 4- or 8-connected graph structures, the cheapest path is simply reconstructed by starting at the initial node and moving towards the cheapest neighbouring node until the goal node is reached. With Field D* the optimal cell boundary point has to be recomputed in each step, as only the costs for the grid corner points are stored in the cost map. In the vicinity of obstacles, the linear interpolation assumption is violated, which leads to errors during path extraction. To reduce those errors, a one-step lookahead is used when computing the next waypoint. Otte and Grudic [5] address another suboptimal phenomenon that occurs during standard path extraction. Paths tend to drift sideways toward their horizontal or vertical bounds, resulting in increased path lengths and unnecessary turns. By using an advanced gradient interpolation technique, the extracted path is significantly improved. However, the resulting path on a uniform obstacle free map still does not correspond to the optimal straight line path between the start and goal point.

Ferguson and Stentz [1] state that the most efficient way to utilize Field D* is to compute the cost-to-goal value function over the grid, and use a local planner to control the actual vehicle trajectory. This means, that the optimal path itself is not extracted. The path planning algorithm operating on NASA's Mars Exploration Rovers (MER) is employing Field D* in this way [2]. A set of candidate arcs starting from the rover's current position are evaluated. Each arc is getting votes based on three criteria: traversability of the traversed cells, required steering action, and the estimated remaining travel cost at the end of the arc. The latter is based on the cost field generated by Field D*. Finally, the arc with most votes is selected for execution and the rover drives a short distance along the arc. This process is then repeated until the goal is reached.

State Lattice

It is not guaranteed that a rover can follow a grid-based path produced with D* or Field D*. The path may include sharp turns, which cannot be executed by mobility-constrained vehicles. In many applications a local planner, which satisfies the differential constraints of the rover, is used in a second step to follow the global path approximately. Another approach is to use a path tracking controller to smooth out the corners. However, in dense environments such techniques can fail when the smoothed path intersects with obstacles, which leads to collisions.

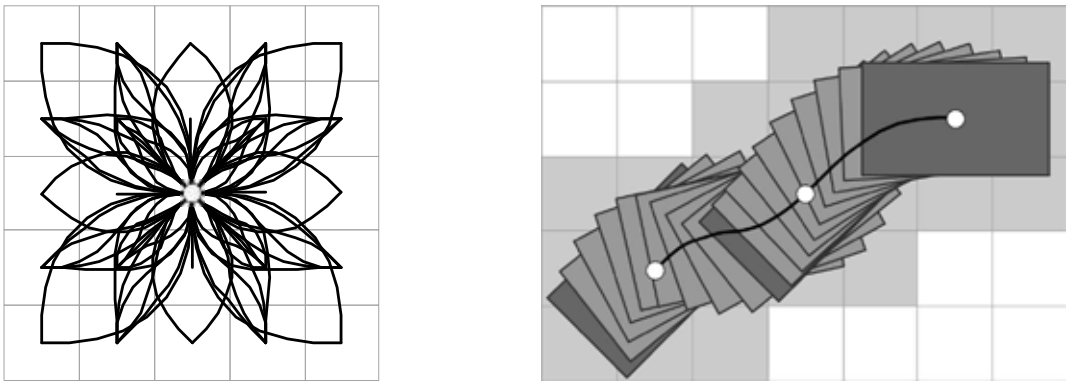


Fig. 1: The State Lattice control set with 8 discrete headings that was designed for the comparison tests (left). The touched grid cells are determined by the swath, which is shown for two consecutive transitions (right).

To overcome this problem, Pivtoraiko et al. [3] present the State Lattice approach that satisfies the differential constraints of the vehicle at planning time. State Lattice employs the D* search algorithm on a grid-based map, where the different cells are connected in such a way, that the search graph satisfies the differential constraints by construction.

The connectivity of the grid cells is defined by the *search space*, which is a vehicle-specific discretization of the robot state. A trajectory generator creates feasible transitions, by considering the motion constraints of the vehicle, such as the minimum turn radius. Furthermore, continuity constraints are imposed, which is in particular the requirement of a continuous heading between two transitions. To enforce this continuity, an additional dimension is added to the search space, so that $(x, y, north)$ and $(x, y, east)$ become distinct states. The set of feasible motions forms the *control set*. As the discretization of the states is performed on a regular pattern, the control set is translational invariant, and can be precomputed. Additionally, the differential constraints of the vehicle have no longer to be considered during planning explicitly, as the search space only consists of motions that satisfy them from the outset.

The cost of a graph edge is calculated using the *swath* of the move. As the state lattice no longer only connects adjacent cells, a single transition can pass over several grid cells, as illustrated in Fig. 1. The cost of a move is computed as the average cost of the touched grid cells multiplied with the length of the path segment. The affected cells of a particular move are precomputed for all transitions in the control set. As the actual dimensions of the rover are considered during the creation of the swath, this method produces more accurate results than expanding obstacles in the map, especially for non-circular vehicles.

Good heuristic cost estimates speed up graph search significantly. However, the euclidean distance as the simplest heuristic estimate is not a well-informed heuristic for differentially constrained vehicles. Especially for short lateral movements, the true path length is underestimated, as the differential constraints require extensive manoeuvring to reach the goal. By exploiting the regularity of the State Lattice, the cost of all possible state pairs is precomputed on an obstacle free map using the planner itself. Those costs are stored in a heuristic look-up table (HLUT) and subsequently used as heuristic estimates to enhance the efficiency of the State Lattice approach.

Most exploratory missions require the rover to approach a point of interest from a specific direction. As the state space is extended by the heading dimension, the orientation of the rover at the goal point is taken into account during path planning. With general D* or Field D* path planners, the rover approaches the goal point from any direction with a non-specified final orientation. Thus, turn-in-place manoeuvres may be required to adjust the rover's orientation at the goal point, which might be costly or even infeasible due to the proximity of obstacles.

If the rover is capable of turning in place, such motions can be included in the control set. This can shorten some complex paths substantially by avoiding tortuous manoeuvring. By tuning the cost of turn-in-place motions, the robot's preference for performing turn-in-place motions versus smooth manoeuvres can be adjusted. Similarly, reverse movements can be added to the control set. This facilitates for example the task of moving one cell sideways with the same start and goal heading. First, the rover moves forward and one cell sideways, before driving backwards to the goal point. Depending on the turn-in-place cost factor, the rover might alternatively perform two 90 degree turns on spot with a straight transition in between.

Obviously, incorporating the differential constraints increases the computational complexity of planning. The extra heading dimension of the state space increases the outdegree of every node. On the other hand, the State Lattice only provides paths that are feasible and thus no local planner is required. This extra effort to ensure an executable path makes the lattice planner about one order of magnitude slower than a grid planner, as stated in [3].

Implementation

In this work, the State Lattice path planner was implemented for an all-wheel steered rocker-bogie rover. For this kind of rover, no minimal turn radii apply and turn-in-place motions are possible. Therefore, the differential constraints used in the State Lattice do not limit the connectivity of the search space. In order to compare State Lattice with other approaches regarding smoothness and computational complexity, the control set was defined manually, pretending there were differential constraints. Eight discrete headings were introduced. The chosen connectivity between the grid cells is a compromise between the ability to express all possible motions and the intention to keep the control set small, in order to minimize the planning complexity. Fig. 1 shows the resulting control set after generating the trajectories using polynomial spirals as described in [6]. Adding turn-in-place motions to the control set ensures that this capability of the all-wheel steered rover can be exploited. Including reverse motions, the designed control set consists of 200 transitions. Due to the continuous heading constraint, only an eighth of the control set is examined at a time, which results in an average outdegree of 25 for a single node.

Rover adapted D* (raD*)

The State Lattice approach produces feasible paths for differentially constrained vehicles, such as Ackerman-steered cars. However, limiting the mobility of an all-wheel steered rover is undesirable, as not all of the mechanical capabilities will be exploited. Inspired by State Lattice, a new path planning approach is presented, which produces smooth paths without restricting the rover's motion capabilities. As this path planner is designed to match the needs of planetary rover missions, it will be referred as *rover adapted D** (raD*).

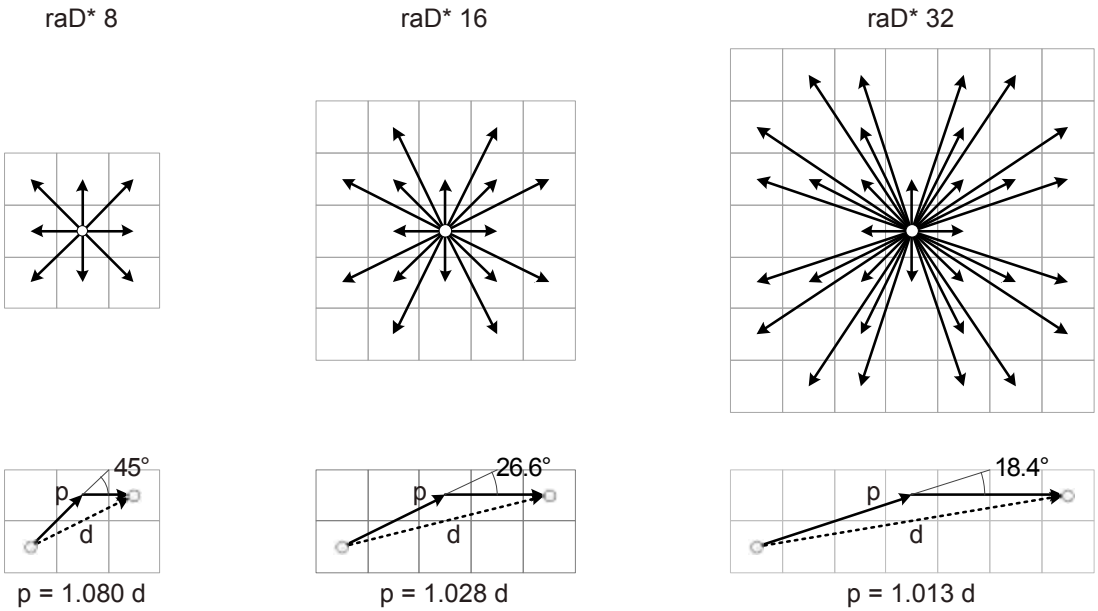


Fig. 2: The top row shows the control sets of the three different versions of raD. The bottom row illustrates the worst-case scenario regarding the excess length of a grid-based "optimal" path p compared to the direct straight line d. The required additional turn is also indicated.*

Three beneficial concepts of State Lattice are adopted. First, the cost calculation is based on the swath of the motion. Second, the desired goal heading is taken into account. Finally, heading continuity is no longer enforced, but favoured using a penalty cost for sharp turns.

As there are no constraints on heading continuity, the search space is no longer extended by an extra heading dimension, for which reason the search space remains two dimensional. To increase the number of possible path headings, the number of connected grid cells was increased. Fig. 2 shows the designed control set for 8, 16, and 32 connected nodes. It also illustrates the worst-case scenario for each control set regarding the "optimal" grid-based path compared to the optimal straight line. The path exceeds the straight-line distance by 8%, 3%, and 1.3% at most, for an 8-, 16-, and 32-connected graph, respectively. The "optimal" paths also include imposed turns of 45°, 26.6°, and 18.4°, respectively.

The edge costs are computed using the swath of the motion for two reasons. First, it generates more accurate results than the commonly used configuration space, where obstacles are expanded by the rover size. Especially for elongated vehicles, the swath approach outperforms the configuration space method, as the exact dimensions of the vehicles are taken into account. Secondly, with 16- or 32-connected graph structures, non-adjacent nodes are connected together. The traversability of the interjacent cells has to be incorporated, which is efficiently done using the precomputed swath of the motion.

During an exploratory mission, the rover is normally directed to approach a defined goal in order to observe a certain point of interest or perform an experiment. To complete the task successfully, the camera or the manipulator arm have to be positioned at a given location. As those tools are mounted on the rover body, not only the position of the rover's centre point is important, but also its orientation at the goal point. To adjust the final heading of the rover at the goal point, the rover could turn on spot. However, the area of interest, and thus the goal

point, is usually close to an obstacle, where a turn-in-place motion is costly or even infeasible. It is therefore desirable to approach the goal with a heading that corresponds to the given final orientation. To ensure this, only paths that approach the goal with the desired heading are considered during planning. All nodes that are connected to the goal node through a path with a heading outside of the defined range are assigned an infinite cost. Similarly, the first path segment should point in the direction of the current rover heading, to prevent costly turn-in-place motions. Therefore, only paths leaving the start node within the given heading range are considered.

As mentioned previously, grid-based paths planned with D* or Field D* may include sharp, instantaneous turns. An all-wheel steered rover can theoretically follow any given path by using its turn-in-place capabilities to align to the next path segment before following it. However, this results in an undesirable stop an go motion. The repeated cycle of deceleration to a full stop, wheel alignment, turning, wheel realignment, and finally acceleration of the rover is moreover highly energy inefficient. Therefore, a path tracking controller is commonly used to smooth the path and to follow it approximately. By tuning the control parameters accordingly, the maximal deviation from the original path can be limited and thus collisions with obstacles can be avoided.

Even though a path tracking controller can smooth a path, unnecessary turns should already be avoided during path planning. Assuming a uniform cost map and a 4-connected graph structure, a goal point that is situated diagonally away from the start point can be reached on different, equally optimal paths. One solution is one long horizontal segment and one long vertical segment with a single 90° turn, as illustrated in Fig. 3. A zigzag pattern with several short horizontal and vertical segments results in the same path length, but includes much more 90° turns. It is not clear, which path should be favoured. The zigzag path remains closer to the real world optimal solution, which is a simple diagonal line. Using a path tracking controller, the repetitive heading changes can be smoothed to approach the straight line solution. However, to avoid a suboptimal wiggly line, the path tracking controller has to be tuned in such a relaxed way, that collision free motions can no longer be guaranteed. Therefore, the first path with less turns should be favoured. While this theoretical example shows that reducing the amount of turns in a real world scenario is desirable, raD*8 with only 8 connected nodes would already result in the optimal diagonal solution.

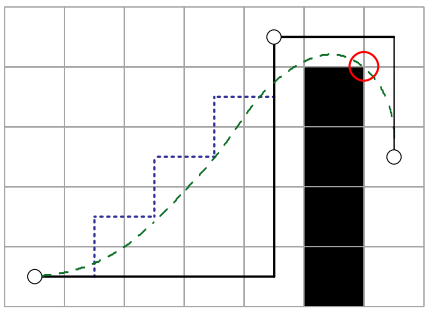


Fig. 3: Two paths with the same length connecting the start and goal points. The solid black path includes less turns than the dotted blue zigzag path. To approach the straight line solution, the path tracking controller has to be tuned in such a way, that the executed path (dashed green line) intersects an obstacle (red circle).

To reduce the amount of turns, a penalty cost p for every heading change is introduced, where ψ_i is the incoming heading angle and ψ_o the outgoing heading angle.

$$p = c_{turn} \cdot |\psi_i - \psi_o|^2 \tag{1}$$

The factor c_{turn} is used to tune the turn penalty cost relative to the traversal cost of the cells. While small turns can be executed without decelerating the rover, a large heading change might require the rover to slow down. Therefore, a path with two turns with an intermediate straight segment is easier to follow than a path with a single turn with the same total amount of heading change. To incorporate this, the difference between the incoming and outgoing headings is squared.

Results

Field D*, State Lattice and raD* were tested and compared. Three versions of raD* were analysed, differing in the number of nodes connected (8, 16, and 32). The turn penalty cost factor of raD* was set to $c_{turn} = 1$. Different maps, and a wide range of distances between the start and goal nodes were used to design 68 search queries. As each query was executed with all five path planning algorithms, 340 paths were analysed for this evaluation. The resolution of the grid map was 0.25 m per cell for a rover that is 0.4 m wide and 0.6 m long. Fig. 4 shows the generated paths of one query for all five tested path planning algorithms.

As all approaches are based on the D* search algorithm, they are all guaranteed to find the optimal path with respect to their cost function. However, the definition of the path cost is different for each approach. Field D* computes the travel cost without considering the swath of the motion nor the desired goal heading. Rover adapted D* on the other hand adds additional costs by penalizing turns. The resulting minimal cost or the path length are therefore not suitable measurements for evaluating the performance of the different path planners.

As the rover used for the tests is all-wheeled steered, any path can be followed by it. Therefore the feasibility of a resulting path cannot be rated. A path is feasible as long as it does not intersect with an obstacle, which is guaranteed by the construction of the map, the search algorithm, and an appropriate tuning of a tracking controller, where applicable. Even though instantaneous turns do not make a path infeasible, they are not desired. The smoother a path, the less instantaneous motions of the rover are required. Hence, the smoothness of the executed path is one value to evaluate the performance of a path planner. To quantify how

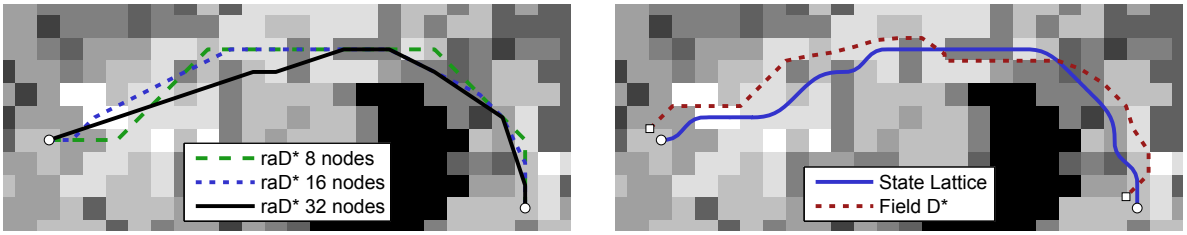


Fig. 4: Resulting paths of one search query, with the paths for the three raD versions in the left map, and the ones of State Lattice and Field D* in the right map. The start and goal nodes of Field D* are slightly shifted, as this algorithm places the nodes on grid cell corners, instead of midpoints.*

smooth a generated path can be followed by the rover, the controller and the kinematics of the rover have to be incorporated.

The actual executed path of the rover was simulated using a path tracking controller based on the Quadratic Curve method [7]. To follow the path, the rover targets a reference point at a certain distance in front of the rover on the path. Compared to other path tracking algorithms, Quadratic Curve has a variable look-ahead distance, based on the curvature of the path. Hence, the rover slows down where the path has narrow turns, which increases the accuracy of the path tracking.

Using a complex kinematic model of the rocker-bogie rover, the required steering angles and rotational speeds of each wheel were then calculated. The derivatives of those values were used to quantify the abruptness of the rover's motion. It was observed, that the maxima of the derivatives are characteristic measurements of the smoothness of a path. The mean and the median of the derivatives showed the same trends. The top row of Fig. 5 shows the

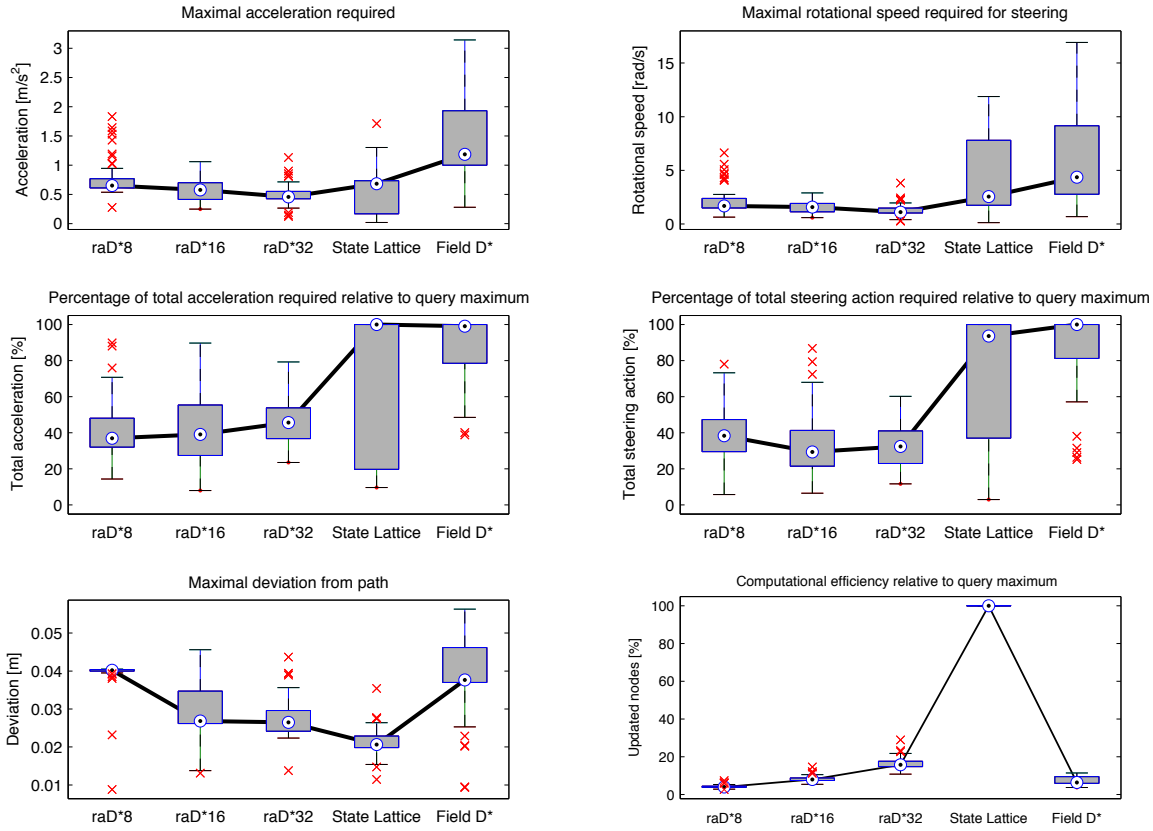


Fig. 5: Results of the performance evaluation for the five path planning algorithms. The results are presented as boxplots, where the circle with the dot represent the median value for the specific algorithm. 50% of the data falls into the grey box, which is bounded by the lower and upper quartiles. The end of the whiskers represent the the lowest and highest data point, that are within 1.5 interquartile ranges (IQR) of the lower and upper quartile, respectively. For normally distributed data, this corresponds to $\pm 2.7\sigma$ or 99.3% coverage. All other values are considered as outliers and are marked with a red cross. To emphasize the trend, the medians of the five algorithms are connected with a black line.

distribution of the maximal derivatives of the required steering angles and rotational speeds for the tested queries. As Field D* does not penalize sharp turns, it was anticipated that this algorithm performs worst with respect to those criteria. For raD*, all three versions perform comparable, as they penalize sharp turns in the same way. However, the outliers imply that raD*8 paths are more likely to have at least one big, instantaneous turn. State Lattice does not perform superior than raD*. This is mainly caused by the path tracking controller applied. The tuning of the controller results in smoothed turns for raD* that are similar to the curves of the State Lattice path segments. Presumably, the performance of State Lattice could be improved using another controller, which uses a priori knowledge about the path segments.

A smooth path is not necessarily superior to another path, if there are a lot of additional turns included. Even though the path can be executed smoothly, every turn involves steering action. In order to quantify this, the total required accelerations and steering actions were summed up. To be comparable, the values in the middle row of Fig. 5 are expressed as percentages with respect to the maximal value of the query among the five algorithms. All raD* versions accumulate significantly less acceleration and steering action than State Lattice and Field D*. Considering that raD*16 and raD*32 generate normally slightly shorter paths than raD*8, they are more energy efficient. State Lattice performs badly regarding steering action, as it involves a lot of turns, especially while driving in a direction which does not correspond to one of the discretized headings. Again, regarding the acceleration required, another controller could presumably improve the performance of State Lattice. Its large variance indicates, that this evaluation for State Lattice is not very significant. Field D* with its abrupt heading changes requires most acceleration and steering action, with a much smaller variance than State Lattice.

The bottom left graph of Fig. 5 shows the maximum deviation from the generated path using the Quadratic Curve controller. As this controller tends to take a short cut in narrow turns, the deviation remains smaller for path planners creating smooth paths. However, by tuning the controller appropriately, the deviation can be kept small for all paths. Further uncertainty of the actual position of the rover relative to the generated path will be introduced by localisation errors. Those localisation disturbances are about the same order of magnitude as the maximal deviations. To avoid collisions, the swath of the rover should be generated slightly larger than the actual rover dimensions, so that localisation errors and deviations from the path can be compensated.

Although raD* was not designed to be faster than the other algorithms, the computational efficiency is still an important factor to evaluate a path planning algorithm. As not all implementations were extensively optimized, the actual running time is not compared. Irrespective of implementation details, the number of nodes that are inserted and updated during the search are a good indicator of the computational complexity of an algorithm. The bottom right graph of Fig. 5 shows the percentage of updated nodes relative to the maximal value of the query among the five algorithms. Field D* and raD* with 8 connected nodes are computationally the least expensive algorithms. With 16 and 32 connected nodes for raD*, the number of updated nodes is approximately doubled and quadrupled, respectively, which correlates to their connectivity. State Lattice expands more nodes than the other algorithms, due to the additional dimension of the search space. This corresponds with the results in [3], where it is noted, that State Lattice is an order of magnitude slower than a grid planner with 16 connected nodes.

Discussion

The results of the comparison tests show, that Field D* produces paths very efficiently. However, extracting paths from its cost field is computationally expensive and erroneous. Even if the mentioned problems of path extracting are diminished with advanced algorithms, the resulting paths consist of sharp instantaneous turns. Although it is possible to follow the path in a more or less smooth way using an appropriate path tracking controller, excessive accelerations and steering actions are required, resulting in an energy inefficient execution of the path. Field D* is therefore not recommended as a unique global path planner. However, if it is combined with a local arc planner, path extraction becomes unnecessary. This combined approach has proven to be successful, for example on NASA's Mars Exploration Rovers.

Using only a single path planner that creates a global plan with sufficient local accuracy, reduces the computational requirements at execution time. State Lattice and raD* are both designed to perform this task. State Lattice creates smooth paths for differentially constrained vehicles, such as Ackerman steered vehicles. In order to design a control set, the mobility of the all-wheel steered rover was artificially constrained. However, the created paths were not smoother to follow, while being an order of magnitude computational more expensive. Moreover, the resulting paths consist of many turns, which add to the required steering actions leading to higher energy consumption. Wiggly paths could be avoided by enlarging the search space, so that the control set connects cells further away with path segments of small curvature. However, this would increase the computational complexity even more.

The presented rover adapted D* (raD*) path planning algorithm creates paths, that are superior to the ones of State Lattice and Field D* with respect to all analysed criteria. Of the three tested raD* versions, differing in the number of connected nodes (8, 16, and 32), the version with the highest outdegree (raD*32) preformed best. Therefore, raD*32 is the recommended path planning algorithm for all-wheeled steered rovers. If computational limitations apply, raD*16 is still an appropriate solution, as it generates paths similar to raD*32, but is computational less expensive.

Conclusion

A new rover adapted path planning algorithm raD* is presented and compared with Field D* and State Lattice for an all-wheel steered rover. Different test scenarios have shown, that Field D* should only be used in combination with a local path planner. Furthermore, it was concluded, that it is not beneficial to artificially constrain the rover's motion capabilities in order to produce a smooth path using State Lattice. It was shown, that raD* produces paths that are feasible, and smoothly trackable more efficiently than State Lattice. Of the three tested versions of raD*, either raD*16 or raD*32 is recommended, depending on the rovers computational resources.

Acknowledgments

This work was performed at the Australian Centre for Field Robotics (ACFR) at the University of Sydney. I would like to thank Dr. Salah Sukkarieh and Dr. Ali Haydar Göktoğan for supervising my project progress.

References

[1] D. Ferguson and A. Stentz, "Using Interpolation to Improve Path Planning: The Field D* Algorithm," *Journal of Field Robotics*, vol. 23, no. 2, pp. 79–101, 2006.

[2] J. Carsten, A. Rankin, D. Ferguson, and A. Stentz, "Global Path Planning on Board the Mars Exploration Rovers," in *2007 IEEE Aerospace Conference*, 2007, pp. 1–11.

[3] M. Pivtoraiko, R. Knepper, and A. Kelly, "Differentially Constrained Mobile Robot Motion Planning in State Lattices," *Journal of Field Robotics*, vol. 26, no. 3, pp. 308–333, 2009.

[4] S. Koenig and M. Likhachev, "D* Lite," in *Eighteenth National Conference on Artificial Intelligence*. American Association for Artificial Intelligence, 2002.

[5] M. Otte and G. Grudic, "Extracting Paths from Fields Built with Linear Interpolation," in *Proceedings of the 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE Press, 2009.

[6] A. Kelly and B. Nagy, "Reactive Nonholonomic Trajectory Generation via Parametric Optimal Control," *The International Journal of Robotics Research*, vol. 22, no. 7-8, p. 583, 2003.

[7] K. Yoshizawa, H. Hashimoto, M. Wada, and S. Mori, "Path Tracking Control of Mobile Robots Using a Quadratic Curve," in *Proceedings of the 1996 IEEE Intelligent Vehicles Symposium*, 1996, pp. 58–63.

Autonomous soaring for atmospheric exploration of Titan

Nicholas R.J. Lawrance and Salah Sukkarieh

Australian Centre for Field Robotics, School of Aerospace, Mechanical and Mechatronic Engineering, University of Sydney, NSW, Australia, 2006

Summary: Unmanned Aerial Vehicles (UAVs) provide a number of advantages for exploration of bodies in the solar system with atmospheres dense enough to sustain aerodynamic flight. However, flight usually requires significant energy input to provide propulsion; requiring additional mass and limiting mission duration. The current work aims to reduce this energy requirement through autonomous soaring. In particular, Titan presents an attractive target for aerial exploration due to favourable gliding conditions and scientific interest. We present a baseline design and navigation algorithm for a gliding fixed-wing aircraft. Results demonstrate the ability of a soaring controller to maintain continuous unpowered flight in conditions expected to be found on Titan.

Keywords: Planetary exploration, Autonomous soaring, Unmanned Aerial Vehicles

Introduction

Exploration of extra-terrestrial bodies in the solar system provides information for understanding the origins of Earth. Furthermore, bodies with a significant atmosphere present a unique opportunity for the use of aerial vehicles for exploration. Aerial vehicles could provide a number of advantages over surface or space-based observers including higher speed, range, and controllability. A number of platform options for aerial vehicles have been considered for aerial exploration including fixed-wing aircraft, balloons, blimps and helicopters [1]. While fixed-wing aircraft are the most common and technologically developed aircraft for use on Earth, they normally require power to overcome drag and provide propulsion. This study proposes the use of soaring flight to reduce the requirement for propulsive power and extend mission duration and range without requiring additional stored energy.

Soaring is the capture of energy from the atmosphere and in favourable conditions can provide continuous flight without propulsive power. Soaring flight is often broken into two techniques; static soaring, which is the use of rising air such as thermal columns, and dynamic soaring, which is the use of spatial distributions of wind speed such as the planetary boundary layer.

Early research into soaring primarily focused on static soaring and techniques for manned glider soaring are well-developed in theory and practice [2]. These theories have been extended into autonomous aircraft with demonstration of autonomous static soaring to compete with remotely-piloted gliders [3, 4]. Early dynamic soaring research focused on off-line optimisation techniques to determine the types of conditions required for soaring [5, 6]. More recent research has examined the possibility of on-line control for dynamic soaring using wind gradients [7] and turbulence [8].

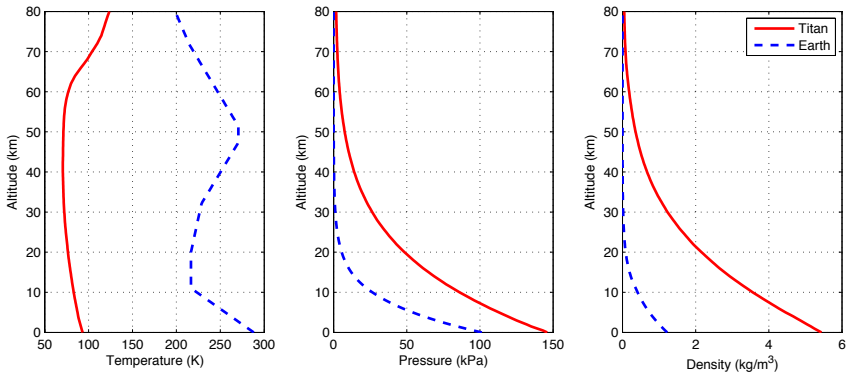


Fig. 1: Properties of the Titan and Earth atmospheres at up to 80km altitude.

While there are a limited number of bodies in the Solar System with sufficient atmosphere for flight (namely Venus, Mars and the Saturnian moon Titan), these targets are naturally highly scientifically interesting. Atmospheric exploration has been considered for all three, with a well-developed concept for a Mars aircraft currently being evaluated by NASA [9]. However, Titan possibly presents the most attractive target for atmospheric exploration due to the favourable conditions for flight (dense atmosphere and relatively low gravity).

Atmosphere of Titan

The atmospheric properties of Titan are an important consideration for the design of a gliding aircraft. Titan's atmosphere is primarily composed of nitrogen, with significant amounts of methane and trace amounts of argon [10]. The methane component forms clouds and rain in the Titan atmosphere as well as a mid-altitude ($40km$) haze which obstructs space-based observation of the surface in the visual spectrum [11].

The atmosphere of the Saturninan moon of Titan is uniquely suited to exploration by an autonomous aircraft. The surface air density and pressure are higher than that of Earth, at $5.44kg/m^3$ and $146kPa$ respectively [12]. The acceleration due to gravity is also lower at $1.354ms^{-2}$ [13]. These conditions are favourable for gliding flight as the reduced gravity requires less lift and the thick atmosphere provides high aerodynamic force at low speed.

Figure 1 illustrates the temperature, pressure and density altitude profiles from the Titan atmosphere engineering model generated by Yelle *et al.* [12] to an altitude of 80km. This is compared against a standard Earth atmosphere model commonly used in aviation [14].

Soaring flight requires wind motion for energy capture. Vertical air motion is necessary for static soaring. Vertical winds on Earth are driven by atmospheric mixing and deflection of lateral winds by ground obstacles. A major energy source for static soaring on Earth is the convective mixing of the lower troposphere driven by solar irradiance of the surface [15]. However, the irradiance on Titan is significantly lower than that on Earth due to the increased distance from the sun and methane haze interference. The presence of clouds observed in the polar regions of Titan suggest that convective mixing is occurring, with estimated vertical air motion between 1.2 and 10 m/s [16]. Recent studies have suggested that polar rain observed on Titan would require vertical air motion of up to $20m/s$ [11]. Of further interest is the

possibility of mountain wind deflection resulting in predictable and useful standing waves of wind. Mountain ranges up to 1.9km high with slopes up to $38°$ have been observed on Titan. Based on these observations and estimated wind profiles, [17] examined the possibility of mountain ranges as a driver for cloud formation. Simulated results suggest that smaller mountains of $500m$ in dry (45% humidity) conditions result in vertical wind speeds around $0.3m/s$ which is insufficient for triggering covective cloud formation but may be enough for low-speed soaring.

Lateral winds are significant at mid-altitudes of the Titan atmosphere. Data collected during the Huygens' descent generated a wind profile with low surface winds (0.1-1.0 m/s) increasing up to $20m/s$ at the $40km$ tropopause [18]. Steady winds are not useful for energy collection (nor do they hamper it) but wind distributions may be. Sufficient wind gradients may be useful energy sources for dynamic soaring but this was not considered in detail in the current study.

Soaring Dynamics

This section presents a brief explanation of soaring and the equations of motion that determine soaring performance. These are used by the soaring controller to estimate energy-gain paths in a wind field. A more comprehensive explanation can be found in the authors' previous work [7].

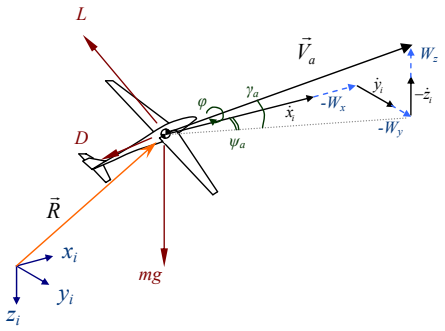


Fig. 2: Air-relative velocity and applied forces for a gliding aircraft.

The glider dynamic model is a six degree-of-freedom (6DOF) flight model for an unpowered aircraft in wind. The applied forces are the aerodynamic force (decomposed into lift, $\vec{L}$, and drag, $\vec{D}$) and weight ($m\vec{g}$). The aerodynamic force is a function of the motion of the vehicle relative to the surrounding air and the physical properties of the aircraft (shape, size, surface properties). Weight is the force due to gravity and is directed down in inertial space. Thus there are two critical frames of reference; an inertial frame which is fixed with respect to the ground, and the air-relative frame which is the aircraft motion relative to the surrounding air. The definition of the air relative climb (γ_a) and heading (ψ_a) angles and the velocity vectors are shown in Fig. 2. The bank angle (ϕ) is the rotation of the lift vector around the velocity vector.

The air-relative velocity vector, $\vec{V}_a$, is the difference between the inertial velocity, $\dot{\vec{R}}$, and the wind component, $\vec{W}$, expressed in the inertial frame. Lift and drag are related through the common approximation of drag coefficient (C_D) as the sum of parasitic ($C_{D,0}$) and lift-induced ($C_{D,i}$) components. The induced drag is a function of the lift coefficient C_L, wing aspect ratio $\mathcal{R}$ and Oswald's efficiency factor e [19]:

$$C_D = C_{D,0} + \frac{C_L^2}{\pi \mathcal{R} e} \tag{1}$$

Wind is considered as a field with locally linear spatial gradients, described by $\mathbf{J_w}$:

$$\mathbf{J_w} = \begin{bmatrix} \frac{\partial W_x}{\partial x} & \frac{\partial W_x}{\partial y} & \frac{\partial W_x}{\partial z} \\ \frac{\partial W_y}{\partial x} & \frac{\partial W_y}{\partial y} & \frac{\partial W_y}{\partial z} \\ \frac{\partial W_z}{\partial x} & \frac{\partial W_z}{\partial y} & \frac{\partial W_z}{\partial z} \end{bmatrix} \tag{2}$$

Assuming roll rate is directly controlled (as a control input to the system) then there are two remaining variables: pitch rate and lift. These equations can be solved to give the lift required for a certain pitch rate (3) or the pitch rate produced for a given lift constraint (4). The resulting equations for the full 6DOF system are shown in (3–7).

$$L = \frac{m}{\cos\phi}\left(V_a\frac{d\gamma_a}{dt} + g\cos\gamma_a - \begin{bmatrix}\sin\gamma_a\cos\psi_a \\ \sin\gamma_a\sin\psi_a \\ \cos\gamma_a\end{bmatrix}^T \mathbf{J_w}\dot{\vec{R}}\right) \tag{3}$$

$$\frac{d\gamma_a}{dt} = \frac{1}{V_a}\left(\frac{L}{m}\cos\phi - g\cos\gamma_a + \begin{bmatrix}\sin\gamma_a\cos\psi_a \\ \sin\gamma_a\sin\psi_a \\ \cos\gamma_a\end{bmatrix}^T \mathbf{J_w}\dot{\vec{R}}\right) \tag{4}$$

$$D = \frac{1}{2}\rho V_a^2 S C_{D,0} + \frac{L^2}{\frac{1}{2}\rho V_a^2 S \pi \text{Æ} e} \tag{5}$$

$$\frac{dV_a}{dt} = \frac{-D}{m} - g\sin\gamma_a - \begin{bmatrix}\cos\gamma_a\cos\psi_a \\ \cos\gamma_a\sin\psi_a \\ -\sin\gamma_a\end{bmatrix}^T \mathbf{J_w}\dot{\vec{R}} \tag{6}$$

$$\frac{d\psi_a}{dt} = \frac{1}{V_a\cos\gamma_a}\left(\frac{L}{m}\sin\phi + \begin{bmatrix}\sin\psi_a \\ -\cos\psi_a \\ 0\end{bmatrix}^T \mathbf{J_w}\dot{\vec{R}}\right) \tag{7}$$

For simulation of a glider in 6DOF these equations can be integrated numerically with specified roll rate and either lift or pitch rate functions. Lift is limited by maximum lift coefficient and a load factor limit.

These equations can also be used to analyse the mechanisms of energy-gain flight. The flight energy of an aircraft can be described as the sum of gravitational potential and kinetic energy. For an aerodynamic platform, the most important measure of kinetic energy relates to the air-relative speed V_a, as this represents the ability to generate lift.

$$E_a = -mgz_i + \tfrac{1}{2}mV_a^2 \tag{8}$$

Taking the time derivative and substituting (6) yields the overall specific power:

$$\frac{\dot{E}_a}{m} = -V_a\frac{D}{m} - gW_z - V_a\begin{bmatrix}\cos\gamma_a\cos\psi_a \\ \cos\gamma_a\sin\psi_a \\ -\sin\gamma_a\end{bmatrix}^T \mathbf{J_w}\dot{\vec{R}} \tag{9}$$

The first term is the power loss due to drag. For a positive airspeed, this is always an energy loss term. The second term is the static soaring term representing the energy gained or lost from vertical motion of the air relative to the fixed inertial frame. The third term is the dynamic soaring term and represents the energy contribution from wind gradients.

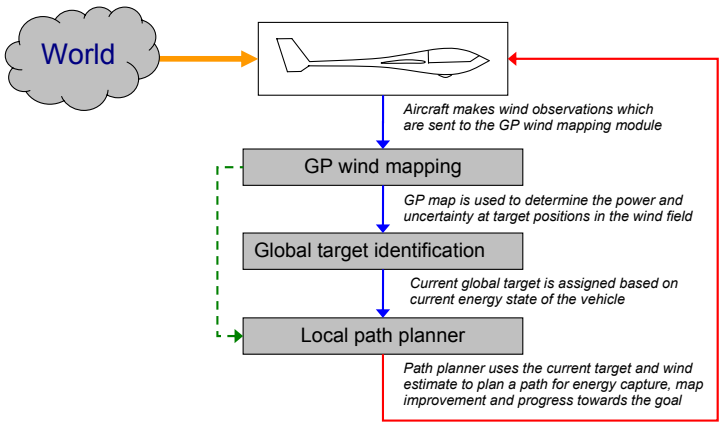


Fig. 3: System overview of simultaneous exploration and exploitation path planning architecture for a soaring UAV.

Simultaneous Exploration and Exploitation of Dynamic Wind Fields

In this paper, the goal is to explore a region of the atmosphere with a gliding aircraft. Exploration is measured using an uncertainty estimate of the current wind map, and is reduced by improving the coverage of samples collected during the flight. However, the gliding platform does not begin with enough energy to reach the entire field, and must collect energy during the flight to allow further exploration.

Previous work by the authors addressed simultaneous exploration and exploitation of a wind field by generating a wind map using Gaussian Process (GP) regression from past observations, then using a hierarchical controller to plan paths through the field to maximise energy gain and information capture [20]. The structure is explained in detail in that work, but an outline is provided here for reference. The overall structure is illustrated in Fig. 3.

The planning structure is divided into three major components; mapping, global target identification and local path planning. Wind observations made during flight are sent to the mapping module, which provides an estimate of the current and future state of the wind field. The global path planner uses the current map and an estimate of the aircraft's current energy level to determine the global target to maximise energy capture and maintain map accuracy. The local planner determines control sequences to maximise energy capture in the current region whilst heading towards the global target. As these plans are carried out, additional observations are made and returned to the mapping module for further replanning.

Gaussian Process wind mapping

Wind estimates are made by taking observations from inertial and air-relative velocities. An air-data system which returns the current airspeed and angles of attack and sideslip can generate a wind vector relative to the vehicle body frame. An inertial system capable of measuring inertial velocity allows observation of the velocity of the vehicle body frame with respect to the inertial frame. Then, the difference between these vectors is the wind vector; the motion of the air relative to the fixed inertial frame. The mapping uses these observations to generate an estimate of the wind field.

Mapping is performed using Gaussian Process (GP) regression, a statistical regression technique. It provides a continuous estimate of both mean and uncertainty of the field from a finite number of observations. The mathematics of GP regression are not covered in detail here but a comprehensive coverage is provided in many other publications, including Rasmussen's work [21] which provided the basis for the GP regression used in this paper. The GP regression model for wind was also presented in the author's previous work [20].

Consider a finite set of observations, $\mathbf{y} = \{y_i\}_{i=1}^n$ where $y_i \in \mathbb{R}$ at locations in the input space $X = \{\mathbf{x_i}\}_{i=1}^n$ where $\mathbf{x_i} \in \mathbb{R}^d$ of an underlying function $f(\mathbf{x})$ with additive zero-mean Gaussian noise with variance σ_n^2. GP regression allows us to draw estimates of the underlying function at any selected test point $\mathbf{x}_*$. This relies on a positive semi-definite covariance function $k(\mathbf{x}, \mathbf{x}')$ which defines the variance between two input points. The estimated mean $\bar{f}_* = \mathbb{E}[f(X_*)|X, y, X_*]$ and covariance $\mathrm{cov}(f_*)$, for covariance function k, covariance matrix $K = K(X, X)$ and target points X_* are shown in (10) and (11) respectively.

$$\bar{f}_* = K(X_*, X)[K + \sigma_n^2 I]^{-1}\mathbf{y} \tag{10}$$

$$\mathrm{cov}(f_*) = K(X_*, X_*) - K(X_*, X)[K + \sigma_n^2 I]^{-1}K(X, X_*) \tag{11}$$

A suitable covariance function should capture the properties of the output regression. Radial Basis Functions (RBF) are commonly used for spatial regression. These are functions based only on a distance measure in the input space such that $k(\mathbf{x}, \mathbf{x}')$ is a function of r where $r = |\mathbf{x} - \mathbf{x}'|$. This corresponds to the properties of a wind field, as we could assume the field to be homogeneous in the gas properties (for small altitude changes) and the variation in velocity to be independent of axis orientation. For this paper, the covariance function is the common squared exponential RBF (12). It is isotropic (input dimensions are equally weighted), stationary (a function only of relative displacement, not of the absolute values of the input points) and non-degenerate (infinitely differentiable).

$$k(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp\left(\frac{-|\mathbf{x} - \mathbf{x}'|^2}{2l^2}\right) \tag{12}$$

In general, there is a set of variable parameters in the covariance function called hyper-parameters θ, where $\theta = (l, \sigma_f)$ in the square exponential RBF. To achieve a 'good' fit of the data, the hyperparameters must be selected to maximise some measure of model fitness. We use the marginal likelihood $p(y|X)$, which is the likelihood of obtaining the training data from the input points marginalised over the model. Calculating the optimal hyperparameters is known as training and usually involves an optimisation routine attempting to maximise log marginal likelihood through variation of the hyperparameters over a set of training data. This is potentially a slow process as the covariance matrix needs to be recalculated and then inverted each time the hyperparameters are varied.

For wind data, the input dimensions are the Cartesian coordinates of the observation location in inertial space and the output is a three-dimensional wind vector corresponding to the Cartesian projection in inertial space. Since the model is shared across the three dimensions, there is only a single matrix inversion required for the three dimensional output space.

Global target planning

The mission goal in the presented scenarios is to explore a region of the atmosphere with insufficient initial energy. The goal of the global planner is to maintain the balance between exploration and exploitation of the atmosphere by defining target waypoints that maximise either energy capture or reduction in map variance. This is achieved using an energy-based utility function which quantifies the current utility of visiting a particular target within the current wind map. Targets are drawn uniformly from the field and ranked using the energy utility.

The first component of the utility function is an estimate of the amount of excess energy the vehicle would have if it travelled to the target. The excess energy e_{excess} is the current aircraft energy, $e_{current}$, minus the target energy e_{target} and an estimate of the energy required to travel to the target e_{travel}. The travel energy is estimated by assuming direct travel to the target for a distance d_{target} at a fixed estimate of the glide ratio $\left(\frac{L}{D}\right)_{est}$.

$$e_{travel} = mgd_{target}\left(\frac{L}{D}\right)_{est}^{-1} \quad (13)$$

$$e_{excess} = e_{current} - e_{target} - e_{travel} \quad (14)$$

The GP regression returns a normal distribution defined by a mean $(\bar{W})$ and variance (σ_W^2) estimate for the wind field velocity at each target point in the map. This is quantified as an energy with an 'optimistic' estimate of the wind using the mean estimate from the current map plus a multiple of the standard deviation (to a maximum of $2\sigma_W$), weighted by the excess energy e_{excess}. The wind estimate is optimistic in unexplored regions when the aircraft has extra energy and pessimistic when it has low energy. The resulting wind estimate is shown in (15) where e_{max} and e_{min} are the energy limits at the highest and lowest altitude of the exploration region respectively.

$$W_{target} = \bar{W} + 2\sigma_W \frac{e_{excess}}{e_{max} - e_{min}} \quad (15)$$

The wind estimate is converted to a power measure using (9) to generate an estimate of the average power available in the region of the target. Dynamic soaring power is calculated by solving for the optimum orientation which would maximise power from the estimated wind gradient. However, as noted earlier, dynamic soaring power can only be collected through cycles and the average power over a full cycle is less than the maximum power collected during the energy gain phases. To account for this, the power estimate of the dynamic term is reduced to 1/3 of the maximum estimate with a minimum power cut-off to ensure that a full cycle would actually produce positive power, based on previous research by the authors [22].

The target utility is estimated each planning cycle, and the target with the highest utility is defined as the current global target and sent to the local planner.

The global target indicates a region where energy gain flight should be possible, but the planning architecture allows flexibility for the local planner to determine the best path in the region of the target. The weighting of the navigation and energy rewards, γ_E, is an exponential decay function based on distance to the goal, as shown in (20) where $\gamma_{E,min}$ is a minimum energy weighting to ensure efficient paths even at large distances from the goal and $\gamma_{E,100}$ is the weight at $100m$ from the goal. The exploration weighting $\gamma_{explore}$ remains constant during each flight.

$$\gamma_E = \gamma_{E,min} + (1 - \gamma_{E,min}) \exp\left(\frac{|d_{goal}|}{100} \log\left(\frac{\gamma_{E,100} - \gamma_{E,min}}{1 - \gamma_{E,min}}\right)\right) \tag{20}$$

Baseline Titan Glider Design

This section presents a suggested baseline design for a fixed wing gliding UAV for the exploration of Titan's atmosphere. The design is based on three major sources; work on the proposed NASA ARES Mars Airplane mission, physical properties of the Huygens lander and terrestrial manned glider and UAV designs.

Of primary importance is the mass of the aircraft delivered and deployed in Titan's atmosphere. A study of aerial vehicles for exploration of Titan suggested that a Delta III lifter could deliver and deploy an aircraft of approximately 200kg to Titan [1]. The Mars ARES airplane (a powered aircraft designed to fly in the much thinner Martian atmosphere) has a planned (wet) mass of $149kg$, with a maximum limit of $175kg$ for a Delta II 2925 launch vehicle [23]. From an aerodynamic perspective increased mass increases speed and wing loading but can reduce gliding performance in rising air; a heavier vehicle can travel faster and farther for better exploration but requires more powerful atmospheric energy conditions. Considering this, a mass of $200kg$ was selected for the baseline design to provide a better indication of the magnitude of wind speeds required for continuous gliding flight.

Physical limitations of the delivery system also limit the size and volume of the deployed platform. For an aircraft, these effectively translate into limitations on wing span (b) and area (S). For a gliding aircraft, efficiency is generally improved with increased wing aspect ratio ($\mathcal{R} = b^2/S$). The Mars ARES airplane design is stored with wings folded which are deployed once released from the entry capsule. The unfolded wing span is $6.25m$ and wing area $7.0m$ [9]. A similar arrangement with the same wing span was considered for the Titan baseline design to take advantage of maximum wing span for efficiency. The reference area can be considerably reduced due to the higher air density and lower gravity in order to maximise aspect ratio. A total wing area of $4.0m^2$ was selected to provide sufficient lifting area for slower flight resulting in an aspect ratio of 9.77.

Table 1: Comparison of aircraft parameters

Parameter ($Units$)	RC Glider	Manned Glider	Mars ARES	Titan baseline
Mass (kg)	12.5	570	149	200
Wing span (m)	4.32	15.0	6.25	6.25
Wing area (m^2)	0.957	10.7	7.00	4.00
Wing loading (N/m^2)	28.4	523	78.3	67.7
Aspect ratio	19.5	21.1	5.58	9.77
Glide ratio	1:25	1:45	-	1:15

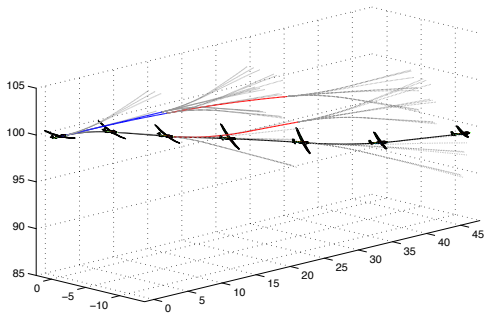


Fig. 4: Local path planning. Paths are estimated for a fixed time horizon with a discrete set of control inputs. Path estimates are ranked with a reward function and the highest ranked paths are further propagated.

Local path planning

The local planner uses the aircraft model, current wind field, and a set of control inputs to plan a path towards the global target whilst maximising local energy capture. A number of alternative controls are considered, and the set of resulting paths are ranked using a local energy utility. Low utility paths are pruned, and the remaining paths are further propagated. The results are ranked again and the process repeated to a specified tree depth. An illustration of the control propagation is shown in Fig. 4.

The local planner reward R is a weighted sum of energy capture, progress towards the global goal and map improvement.

$$R = \gamma_E R_{energy} + (1 - \gamma_E) R_{nav} + \gamma_{explore} R_{explore} \tag{16}$$

The energy capture reward R_{energy} is the energy gained. Previous work [7] found that the performance of this utility is improved using an estimate of the terminal power $\dot{E}_2$ to favour paths that terminate in high power regions for improved performance of following segments.

The navigation reward R_{nav} quantifies the energy advantage of travelling towards the global goal. This is determined by calculating the distance travelled towards the goal, Δd_{goal}, and converting this distance into an energy estimate using the nominal glide ratio.

The exploration reward $R_{explore}$ represents the advantage of exploring new regions of the map. As in the global planner case, this is an energy reward based on an optimistic estimate of the wind. The exploration energy is estimated by taking the variance at each end of the segment and determining the wind within one standard deviation of the mean which would result in maximum energy gain. $\bar{V}$ is the original mean airspeed $\frac{1}{2}(V_1 + V_2)$, and ΔV is the estimated change in airspeed under the new conditions $V_2' - V_2$.

$$R_{energy} = -mg(z_2 - z_l) + \tfrac{1}{2}m\left(V_2^2 - V_1^2\right) + \gamma_{\dot{E}}\dot{E}_2\Delta t \tag{17}$$

$$R_{nav} = mg\Delta d_{goal}/(\tfrac{L}{D})_{est} \tag{18}$$

$$R_{explore} = \tfrac{1}{2}(\sigma_{Wz,1} + \sigma_{Wz,2})\, mg\Delta t + \tfrac{1}{2}m\left(2\bar{V}\Delta V + \Delta V^2\right) \tag{19}$$

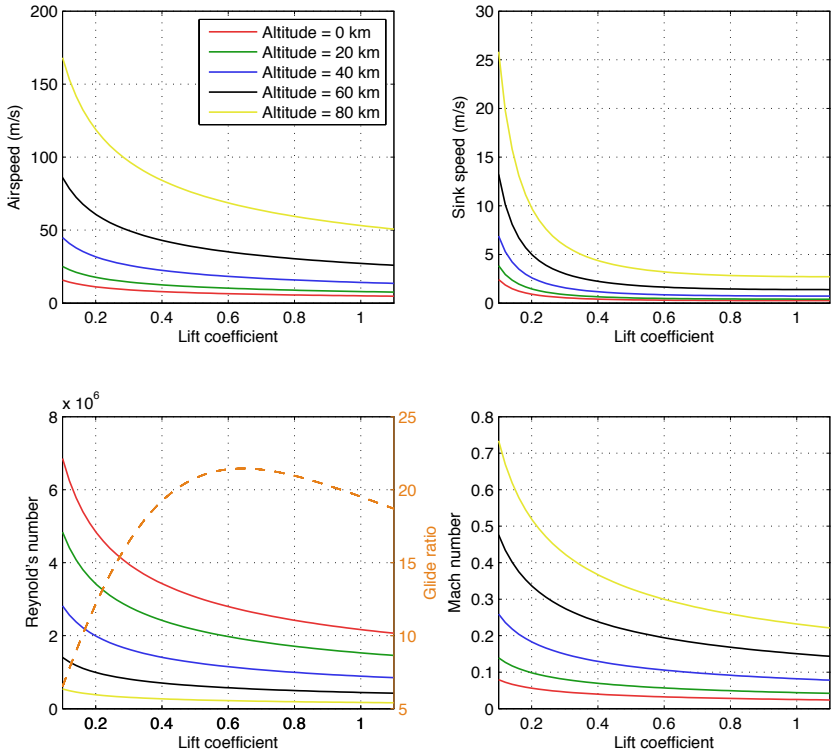


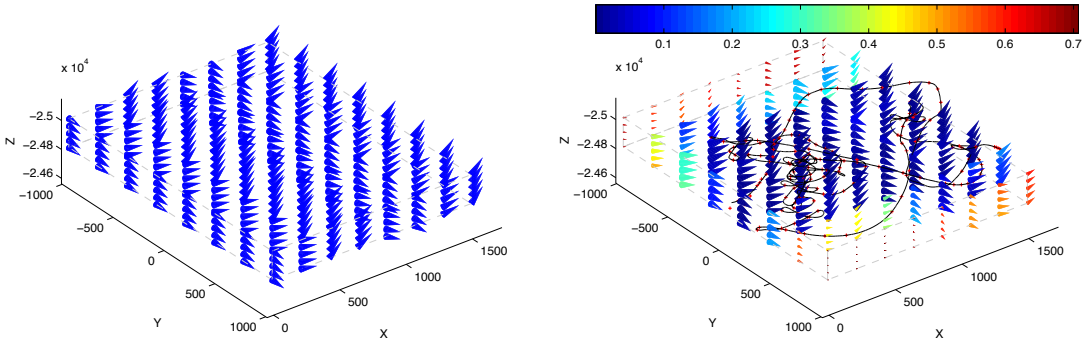
Fig. 5: Flight conditions for Titan baseline design at altitudes between 0 and 80 km in the Titan atmosphere. Glide ratio is independent of altitude.

The Titan baseline design parameters are shown in Table 1. Also shown are values for a Remotely Controlled (RC), manned glider, and the Mars ARES design [23]. The RC glider is an RnR SB-XC as used in previous autonomous soaring research [3, 7]. The manned glider figures are from the flight manual of the DG-808S $15m$ competition grade single seat glider [24]. Note that in all cases the glide ratio is dependent on many factors and the values shown are approximate values at best cruise conditions.

Figure 5 shows the flight conditions for a steady glide (constant airspeed) with the baseline configuration at different altitudes from the Titan atmospheric model [12]. While gliding flight is still theoretically possible at up to $80km$ altitude with the current model, the rapid sink speed would result in significant energy loss. Previous studies have considered flight up to and around $40km$ altitude where there is significant methane haze which would possibly obscure ground observations. Of note is that below $40km$ and with lift coefficients above 0.3, typical values for sink speed are around $0.5m/s$. Thus, relatively low vertical winds would be required for sustained flight of a gliding aircraft. As discussed earlier, winds of this magnitude could be reasonably expected in certain regions of the Titan atmosphere.

Simulation Results

Simulated results are presented to demonstrate autonomous soaring control for a gliding vehicle in the Titan atmosphere. The glider model is the baseline design discussed in the previous section. Wind data is collected by a simulated air data system assuming direct observation of the wind with additional unbiased Gaussian noise with standard deviation



(a) Actual wind field

(b) Final wind map at $t = 2400s$.

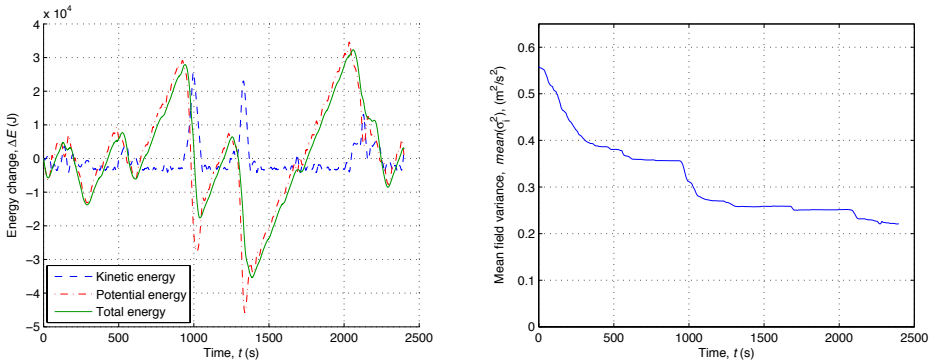
Fig. 6: Actual and estimated wind field. Flight path is indicated by a black line, sample set are red + and the wind speed uncertainty in m/s is indicated by the cone colours.

0.1m/s to represent measurement error. The simulated sensor system collects data at frequency 2Hz but the number of stored training points is fixed at 150 to prevent excessive computational load. The relevance of the data is effectively calculated as part of the covariance function for prediction, so observations that are too close together or too old are removed and replaced with new observations. The set maintenance assumes that only current or future prediction of the wind field are required. This results in a natural spatial sparsity of the data set to provide the best coverage of the target space with the limited number of observations made along the flight trajectory. During the flight the hyperparameters are retrained every 10 seconds with at most 5 conjugate gradient minimisation steps.

The local planner plans for a total horizon of 5 seconds, but only the first 3 seconds of the plan are used before replanning. There are three roll rate and three pitch rate commands for a total of nine control options for each branch. The control sequence returned by the path planner is carried out open-loop by the simulated aircraft with wind data drawn from the simulated wind field.

The wind field is a standing wave represented by a sinusoidal variation of wind along the x-axis with wavelength $1000m$ to represent wave flow generated by a surface obstacle such as a mountain ridge. The wind speed is $2m/s$ with a sinusoidal variation in the vertical direction of up to $\pm 45\,\mathrm{deg}$, resulting in a peak vertical wind speed of $1.41m/s$, as shown in Fig. 6(a). Atmospheric properties are calculated using the Yelle model [12]. Flight begins at an altitude of $24.9km$. The exploration region is defined by a box with axis limits (in metres) $x \in [0, 1500], y \in [-1000, 1000], z \in [24800, 25000]$.

Figure 6(b) shows the flight path and field estimate after $2400s$ of simulation. Also shown in Fig. 7 are the platform energy and map uncertainty over the flight. There are distinct periods of exploration where the map is explored at the cost of energy. When energy decreases too much the controller tends to focus on exploitation by returning to known energy-gain regions of the map and collecting enough energy to continue exploration.



(a) Energy change (b) Mean map uncertainty

Fig. 7: Platform energy and mean map uncertainty during simulated exploration.

Discussion

The goal of this study was to present the potential energy savings which could be expected by utilising autonomous soaring for a fixed-wing aircraft mission on Titan. However, there are a considerable number of limitations to an actual aerial vehicle mission to Titan which are considered beyond the scope of this work. Firstly, physical limitations such as temperature control and (methane) icing are not considered as they are a function of the overall mission and not specifically affected by gliding flight.

The current path planning architecture requires inertial position estimates to generate an inertial wind map. For a vehicle on Titan this is a difficult task as there is no global positioning system available and inertial dead-reckoning estimates will drift over time. One possible solution is to use spatio-temporal GP mapping. This is a relatively trivial extension to the current spatial mapping but assumes that the wind map is a function of time. Since only forward prediction is required, old observations become irrelevant and are discarded, so that the map is continuously updated with new data. This has been considered in concurrent work by the authors [25]. In this case, the inertial drift is less important as the map effectively becomes a local (drifting) wind-relative map which is still useful for energy gain path planning. Updates from external observations such as ground and space communications could provide irregular updates to obtain global position estimates.

Conclusion

We have presented a baseline design and guidance algorithm for an autonomous soaring aircraft for exploration of Titan. We demonstrated that due to the favourable flying conditions on Titan a glider would require approximately $0.5m/s$ of vertical air motion to provide continuous flight. Simulated results show autonomous unpowered exploration of a wind field with zero mean wind and maximum vertical wind speed of $1.4m/s$. Previous research has suggested that such conditions are likely to exist on Titan and could be utilised by an autonomous soaring aircraft with little or no prior information. Use of a soaring aircraft supplemented with powered flight could provide the advantages of fixed-wing aircraft for exploration with a significant reduction in stored energy required for propulsion.

Acknowledgments

This work is supported by the ARC Centre of Excellence programme, funded by the Australian Research Council (ARC), the New South Wales State Government and the Australian Centre for Field Robotics at The University of Sydney.

References

[1] R. Lorenz, "Post-Cassini exploration of Titan: Science rationale and mission concepts," *Journal of the British Interplanetary Society*, vol. 53, no. 7/8, pp. 218–234, 2000.
[2] P. B. MacCready, "Optimum airspeed selector," *Soaring*, vol. January-February, pp. 10–11, 1958.
[3] M. J. Allen, "Guidance and control of an autonomous soaring UAV," NASA Dryden Flight Research Center, Technical Memorandum NASA/TM-2007-214611, February 2007.
[4] D. Edwards, "Implementation details and flight test results of an autonomous soaring controller," in *AIAA Guidance, Navigation and Control Conference*, Honlulu, Hawaii, 2008, AIAA Paper 2008-7244.
[5] Y. J. Zhao, "Optimal patterns of glider dynamic soaring," *Optimal Control Applications and Methods*, vol. 25, no. 2, pp. 67–89, 2004.
[6] G. Sachs, "Minimum shear wind strength required for dynamic soaring of albatrosses," *Ibis*, vol. 147, no. 1, pp. 1–10, 2005.
[7] N. R. Lawrance and S. Sukkarieh, "Wind energy based path planning for a small gliding unmanned aerial vehicle," in *AIAA Guidance, Navigation and Control Conference*, Chicago, Illinois, 2009, AIAA Paper 2009-6112.
[8] J. W. Langelaan, "Gust energy extraction for mini and micro uninhabited aerial vehicles," *Journal of Guidance, Control, and Dynamics*, vol. 32, no. 2, pp. 463–472, 2009.
[9] M. Guynn, M. Croom, S. Smith, R. Parks, and P. Gelhausen, "Evolution of a Mars airplane concept for the ARES Mars scout mission," in *2nd AIAA "Unmanned Unlimited" Systems, Technologies, and Operations - Aerospace, Land, and Sea Conference, Workshop and Exhibition*, 2003, AIAA Paper 2003-6578.
[10] H. Niemann, S. Atreya, S. Bauer, G. Carignan, J. Demick, R. Frost, D. Gautier, J. Haberman, D. Harpold, D. Hunten *et al.*, "The abundances of constituents of Titan's atmosphere from the GCMS instrument on the Huygens probe," *Nature*, vol. 438, no. 7069, pp. 779–784, 2005.
[11] R. Hueso and A. Sánchez-Lavega, "Methane storms on Saturn's moon Titan," *Nature*, vol. 442, no. 7101, pp. 428–431, 2006.
[12] R. Yelle, D. Strobel, E. Lellouch, and D. Gautier, "Engineering models for Titan's atmosphere," *Huygens: Science, Payload and Mission*, pp. 243–256, 1997.
[13] G. Lindal, G. Wood, H. Hotz *et al.*, "The atmosphere of Titan: An analysis of the Voyager 1 radio occultation measurements," *Icarus*, vol. 53, no. 2, pp. 348–363, 1983.
[14] COESA, "US Standard Atmosphere, 1976," *US Government Printing Office, Washington, DC*, 1976.
[15] C. Ahrens, *Meteorology today: an introduction to weather, climate, and the environment*, 9th ed. Brooks/Cole, Cengage Learning, 2008.
[16] M. Awal and J. Lunine, "Moist convective clouds in Titan's atmosphere," *Geophysical Research Letters*, vol. 21, no. 23, pp. 2491–2494, 1994.
[17] E. L. Barth, "Cloud formation along mountain ridges on titan," *Planetary and Space Science*, vol. In Press, Corrected Proof, pp. –, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/B6V6T-50JHBXV-1/2/0ebb68b0d5dc1acd139cd2a274913f14
[18] W. Folkner, S. Asmar, J. Border, G. Franklin, S. Finley, J. Gorelik, D. Johnston, V. Kerzhanovich, S. Lowe, R. Preston *et al.*, "Winds on Titan from ground-based tracking of the Huygens probe," *Journal of Geophysical Research*, vol. 111, no. E7, p. E07S02, 2006.
[19] J. D. Anderson, Jr., *Fundamentals of Aerodynamics*, 4th ed., ser. McGraw-Hill Series in Aeronautical and Aerospace Engineering. New York, New York: McGraw-Hill, 2007.
[20] N. R. Lawrance and S. Sukkarieh, "Simultaneous exploration and exploitation of a wind field for a small gliding uav," in *AIAA Guidance, Navigation and Control Conference*, Toronto, Ontario, Canada, 2010, AIAA Paper 2010-8032.
[21] C. E. Rasmussen and C. K. Williams, *Gaussian Processes for Machine Learning*, ser. Adaptive computation and machine learning. Cambridge, Massachusetts: The MIT Press, 2006.
[22] N. R. Lawrance and S. Sukkarieh, "A guidance and control strategy for dynamic soaring with a gliding uav," in *IEEE International Conference on Robotics and Automation*, Kobe, Japan, 2009, pp. 3632–3637.
[23] R. Braun, H. Wright, M. Croom, J. Levine, and D. Spencer, "Design of the ARES Mars airplane and mission architecture," *Journal of spacecraft and rockets*, vol. 43, no. 5, pp. 1026–1034, 2006.
[24] *Flightmanual for the sailplane DG-808S*, DG Flugzeugbau GmbH, Bruchsal, Germany, 2003.
[25] N. R. Lawrance and S. Sukkarieh, "Path planning for autonomous soaring flight in dynamic wind fields," in *IEEE International Conference on Robotics and Automation*, Shanghai, China, 2011, Submitted for publication.

High Level Risk Analysis and Decision Making Regarding the Prediction of Thermal Lift Locations for an Autonomous Mars Glider

Jen Jen Chung, Salah Sukkarieh

Australian Centre for Field Robotics
School of Aerospace, Mechanical and Mechatronic Engineering
University of Sydney
NSW Australia

Summary: The ARES (Aerial Regional-scale Environmental Survey of Mars) glider project is a current NASA program that aims to send a fully autonomous glider system to survey the Martian landscape from a gliding altitude of 1.5km. The design of the glider propulsion system is complicated by issues of low atmospheric density and the lack of appreciable amounts of O_2 in the Martian environment, thus it is desirable to make use of the glider's soaring capabilities to maximise gliding endurance by exploiting thermal energy sources in the region whilst also considering the external mission goal of exploring the environment. This paper discusses the application of a Bayesian Forecast-Decision System (BFS) to the prediction of thermal motion, the detection and management of system uncertainties, and high level risk assessment of available flight paths through the environment. The results of this paper show that the forecast decision system requires very little a priori information to effectively gauge the system uncertainties and manage platform energy to efficiently survey the allocated search region. Furthermore, it is also shown that the system can run in real-time on an autonomous sensing and reacting system operating in a dynamic environment.

Keywords: Autonomous glider, Bayesian decision making, risk assessment.

Introduction

ARES Scout Mission

The primary science objectives of the ARES Mars Scout Mission are to sample the Martian atmosphere to characterise the structure and dynamics of the atmospheric boundary layer over regional scales, measure water-equivalent hydrogen abundance and ice burial depth, and investigate the crustal magnetism on Mars [1-3]. The mission design requires the ARES glider to explore up to 610km of the Southern Highlands region from a gliding altitude of 1.5km [4]. However, low atmospheric density and the lack of appreciable amounts of O_2 in the Martian environment pose a number of issues to the design of the glider propulsion system; thus it is desirable to make use of the glider's soaring capabilities. To maintain the level of autonomy necessary to successfully carry out the mission, the glider system must be capable of seeking out energy sources within the environment to extend gliding endurance whilst maintaining the nominal pre-planned trajectory to ensure that sufficient data are collected.

The hardware system aboard the ARES glider is shown in Fig. 1. On board sensors are capable of sampling the local atmosphere to determine air temperature and pressure to characterise nearby thermals. Given this information, it is proposed that a Bayesian forecast-decision system can quantify the uncertainties in the sensor measurements and forecast models to plan a path through the environment that accounts for the risks associated with each waypoint decision.

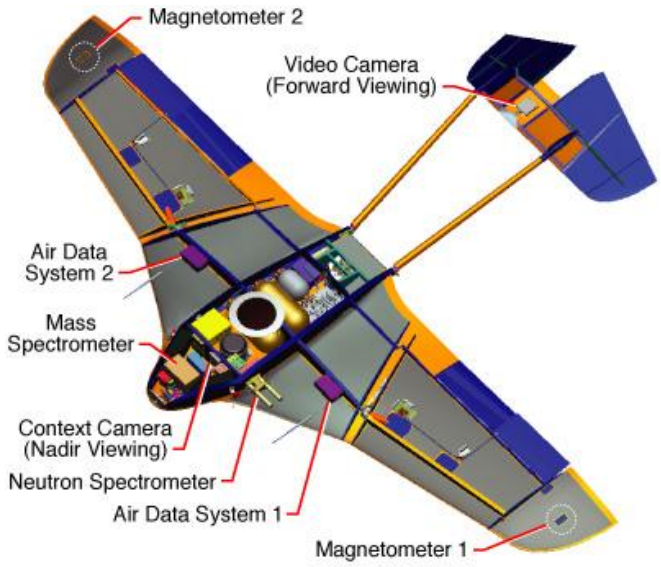


Fig. 1: The NASA ARES platform is equipped with sensors to measure the crustal magnetisation of Mars as well as sample the Martian atmosphere to determine sources of biogenic, volcanic, and chemically active gases [4].

The Bayesian Forecast-Decision System for Time Series Data

The Bayesian forecast-decision system (BFS) was first presented by Krzysztofowicz for the purpose of identifying and quantifying sensor observation and forecast model uncertainties and was applied to meteorology forecasting and sensor fusion problems [5]. The BFS is able to predict thermal motion according to a user defined forecast model, and applies the Bayesian Processor of Forecast (BPF) to detect and manage the contribution of model and input uncertainties associated with glider sensor data and forecast data to produce a posterior distribution over future thermal positions. In general, the glider decision system receives and processes information regarding the state of nature via the timing schematic shown in Fig. 2.

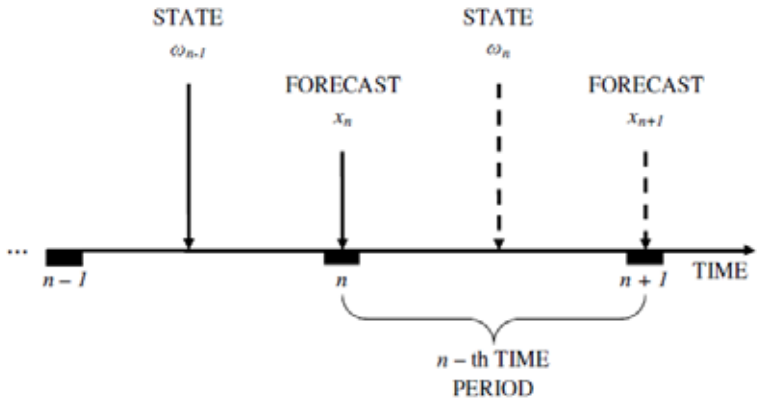


Fig. 2: Schematic timing of information as processed by the BFS. At the beginning of the n-th time period, the system has observations of the previous states and knowledge of all the forecasts including the current forecast, x_n, upon which a manoeuvre decision is made. The state of nature during the n-th time period, ω_n, is observed by the system at the end of this time period and is used to formulate the next forecast x_{n+1}.

The role of the decision model of the BFS is to select the next waypoint by assessing the risk associated with traversing available paths through the environment, where risk is computed according to goal point reachability and the map uncertainty as quantified by the BPF.

Algorithm Design

Forecast Model

The purpose of the initial stage of the BFS is to gather the available raw data and generate a forecast of the state of nature for the next period of time. In this case, a simplified forecasting model is used, which assumes that over relatively short time intervals (within the space of one minute), thermals travel at a constant velocity. Examination of the previous thermal locations can provide a measure and spread of the thermal velocities, however the number of prior states to use should be large enough to account for outliers but also small enough pick up changes in thermal velocities over time. Additionally, an initial estimation of the thermal velocities is required as no prior states have been recorded. The spread of the estimated velocities is later used by the BPF to quantify the uncertainties stemming from this modelling method.

Bayesian Processor of Forecasts

The BPF judges the performance of the forecasting model by comparing previous forecasts against known thermal locations to quantify the current forecast uncertainty. For time series data, the BPF updates the prior distribution of the uncertain state with the new forecast x_n, and employs Bayes' theorem to produce a posterior distribution η_n over the possible future thermal positions ω_n. Namely:

$$\eta_n(\omega_n \mid \omega_1,...,\omega_{n-1},x_1,...,x_n) = \frac{f_n(x_1,...,x_n \mid \omega_1,...,\omega_n)g_n(\omega_1,...,\omega_n)}{\xi_n(\omega_1,...,\omega_{n-1},x_1,...,x_n)} \tag{1}$$

Where ξ_n is the predictive density of the forecast:

$$\xi_n(\omega_w,...,\omega_{n-1},x_1,...,x_n) = \int f_n(x_1,...,x_n \mid \omega_w,...,\omega_n)g_n(\omega_w,...,\omega_n)d\omega_n \tag{2}$$

Densities f_n and g_n are the known joint and prior densities of the relevant variables, respectively. The prior density describes the natural uncertainty regarding the process (in this case, thermal motion) that exists before receiving the forecast. The joint density for fixed forecasts represents the likelihood of the states, that is, the function calculates the forecast uncertainty by judging how well the forecasting model has predicted the state in the past.

Krzysztofowicz [5] suggests a number of assumptions that can be made about the structure of the likelihood function and the relationship of the states to simplify the posterior distribution calculation.

Assumption 1: Conditional upon $\omega_1,...,\omega_n$, forecasts $x_1,...,x_n$ are stochastically independent of each other.

$$f_n(x_1,...,x_n \mid \omega_1,...,\omega_n) = \prod_{i=1}^{n} f_i(x_i \mid \omega_1,...,\omega_n) \tag{3}$$

Assumption 2: Conditional upon ω_i, forecast x_i is stochastically independent of ω_j for all $j \neq i$. Thus, for any $i \leq n$,

Assumption 3: The time series of thermal locations, states $\omega_1,...,\omega_n$, form a first-order Markov process.

$$g_n(\omega_1,...,\omega_n) = \prod_{i=1}^{n} g_i(\omega_i \mid \omega_{i-1}) \tag{5}$$

$g_1(\omega_1 \mid \omega_0) = g_1(\omega_1)$ is the initial state density, and $g_2,...,g_n$ are one-stage transition densities.

Given Eqn 3, Eqn 4 and Eqn 5, Eqn 1 takes the form:

$$\begin{aligned} \eta_n(\omega_n \mid \omega_1,...,\omega_{n-1},x_1,...,x_n) &= \frac{f_n(x_n \mid \omega_n) g_n(\omega_n \mid \omega_{n-1})}{\int f_n(x_n \mid \omega_n) g_n(\omega_n \mid \omega_{n-1}) d\omega_n} \\ &= \eta_n(\omega_n \mid \omega_{n-1}, x_n) \end{aligned} \tag{6}$$

The full derivation can be found in [6]. According to Eqn 6, given the thermal positions during the previous time period and the current forecast for the thermal positions at the next time period, the posterior state transition density $\eta_n(\cdot \mid \omega_{n-1}, x_n)$ over the possible thermal locations in the next time period is a combination of the natural uncertainty about the state and the uncertainty in the forecast x_n of ω_n, the prior state transition density $g_n(\cdot \mid \omega_{n-1})$ and the likelihood function $f_n(x_n \mid \cdot)$, respectively.

The natural uncertainty can be estimated based on on-site historical records, regional information and any scientific knowledge or experience relevant to ω_n. For the purpose of locating thermals, this could include the tendency for thermals to diminish within certain periods of time (particularly if they are likely to disappear altogether in a space of time shorter than the data sampling time), and other known thermal movement patterns and thermal interaction behaviour.

The uncertainty in the forecast x_n of ω_n is estimated from historical records; by observing J joint observations of forecasted thermal locations and actual thermal locations $\{(x_j,\omega_j) : j = 1,...,J\}$, it is possible to determine the performance of the forecasting model. At the start of operation, no adequate historical record of the forecast and actual states will exist. Krzysztofowicz suggests that under this circumstance, a simulated record can be used; however, care must be taken to ensure that the simulated model accounts for all uncertainties and assumptions of the actual forecasting model. As the glider begins to receive sensor observations, the significance of the historical records on the state estimate will decrease as a sliding window is used to regulate the number of observation entries used in the estimation.

Given the estimates of f_n and g_n, the predictive density of the forecast can be calculated as:

$$\xi_n(x_n \mid \omega_{n-1}) = \int f_n(x_n \mid \omega_n) g_n(\omega_n \mid \omega_{n-1}) d\omega_n \tag{7}$$

With $\xi_1(\omega_1 \mid \omega_0) = \xi_1(\omega_1)$. Combining Eqn 6 with Eqn 7 gives:

$$\eta_n(\omega_n \mid \omega_{n-1}, x_n) = \frac{f_n(x_n \mid \omega_n) g_n(\omega_n \mid \omega_{n-1})}{\xi_n(x_n \mid \omega_{n-1})} \tag{8}$$

Fig. 3 shows the layout of the BPF algorithm.

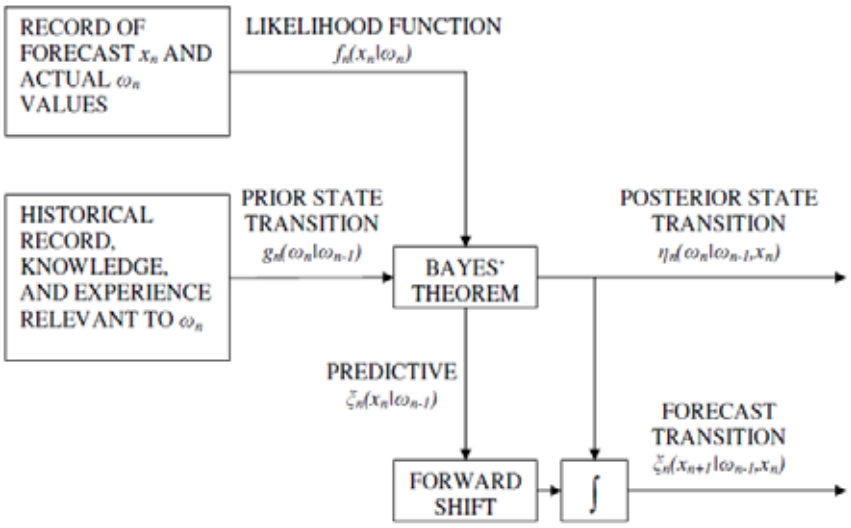


Fig. 3: Simplified block diagram showing the steps involved in the BPF algorithm for a first-order Markov Process [5].

With respect to the overall risk analysis system, the variables of interest from the BPF are the posterior mean and variance of η_n. Since the likelihood and prior densities are both assumed to be normal, then it can be seen that g_n is a conjugate prior of f_n. Thus it is possible to derive a closed form representation of the posterior distribution from which the two hyperparameters of interest can be obtained.

Supposing that the relationship between the forecast x_n and the actual state ω_n can be modelled using a linear forecast equation:

$$x_n = a_n \omega_n + b_n + \theta_n \tag{9}$$

where a_n and b_n are fixed parameters and θ_n represents the residual noise with a normal density distribution $N(0,\zeta_n^2)$. In the case of perfect forecasts, $a_n = 1$ and $b_n = \zeta_n = 0$. Therefore it is expected that the calculated values of these parameters converge to this ideal condition. Computation of the parameters necessitates that b_n and θ_n are measured as one variable with a distribution $N(b_n,\zeta_n^2)$ from which the required parameter values can be extracted.

From Eqn 9, the likelihood function can be represented as a normal distribution:

$$f_n(x_n \mid \omega_n) = N\left(a_n \omega_n + b_n, \zeta_n^2\right) \tag{10}$$

Or in terms of the forecast:

$$\begin{aligned} f_n(x_n \mid \omega_n) &= N\left(\mu_f, \sigma_f^2\right) \\ &= N\left(\frac{x_n}{a_n} - \frac{b_n}{a_n}, \frac{\zeta_n^2}{a_n^2}\right) \end{aligned} \tag{11}$$

In a similar fashion, since the states form a first-order Markov process, then:

$$\omega_n = c_n \omega_{n-1} + d_n + \nu_n \tag{12}$$

where c_n and d_n are fixed parameters and v_n represents the residual noise with a normal density distribution $N(0,\tau_n^2)$. Once again, if perfect thermal data are received from the sensors and the thermals behaved exactly according to the linear motion assumption made in the forecast model, then $c_n = 1$, $d_n = \Delta t \times v$ and $\tau_n = 0$, where $\Delta t \times v$ is the time step multiplied by the thermal velocity. Further, computation of d_n and v_n must also combine the two parameters into a single distribution $N(d_n,\tau_n^2)$ containing both values of interest.

Eqn 12 gives:

$$\begin{aligned} g_n(\omega_n \mid \omega_{n-1}) &= N\left(\mu_g, \sigma_g^2\right) \\ &= N\left(c_n \omega_{n-1} + d_n, \tau_n^2\right) \end{aligned} \tag{13}$$

Applying Eqn 11 and Eqn 13 into Eqn 7 and Eqn 8, the mean and variance of the posterior distribution can be calculated using the posterior hyperparameter relations derived by Raiffa and Schlaifer [7]. That is, for:

$$\eta_n = N\left(\mu_n, \sigma_n^2\right) \tag{14}$$

$$\begin{aligned} \mu_n &= \frac{\left(\frac{\mu_g}{\sigma_g^2} + \frac{\mu_f}{\sigma_f^2}\right)}{\left(\frac{1}{\sigma_g^2} + \frac{1}{\sigma_f^2}\right)} \\ &= \frac{a_n \tau_n^2 x_n + c_n \zeta_n^2 \omega_{n-1} + d_n \zeta_n^2 - a_n b_n \tau_n^2}{\zeta_n^2 + a_n^2 \tau_n^2} \end{aligned} \tag{15}$$

$$\begin{aligned} \sigma_n^2 &= \left(\frac{1}{\sigma_g^2} + \frac{1}{\sigma_f^2}\right)^{-1} \\ &= \frac{\tau_n^2 \zeta_n^2}{\zeta_n^2 + a_n^2 \tau_n^2} \end{aligned} \tag{16}$$

Decision Model

The purpose of the decision model is to allocate a discrete action to any given state of nature. In the BFS, the decision model receives the probability distributions calculated in the BPF to create a map of the world including the glider's own position and where it expects thermals to be located. Given the estimated state and the respective probabilities associated with each thermal's existence, the model is able to choose a suitable target for the glider. The decision of which option represents the 'best' choice is based on the calculated expected utility resulting from each manoeuvre.

For the glider, utility is determined as a weighted combination of three factors:

- Distance of detour as a fraction of the nominal path
- Deviation of glider heading angle from nominal path
- Energy/altitude gain available

The weighting is an operator defined heuristic and should reflect the mission objectives. For the given exploration and survey missions outlined in this paper, a weighting ratio of 20:1:10 is applied, respectively. The low weighting for the glider heading angle deviation reflects the manoeuvrability of the platform while the distance factor is weighted highest to drive the glider to its goal location.

In addition to these factors, the decision model is also able to safely direct the glider when it is operating below an operator defined critical altitude. Under such circumstances, the decision model instructs the glider to apply powered flight to the nearest sensed thermal.

Glider Simulation Design

The glider simulation is set up such that for each trial, a random set of thermals are generated in the field of operation. Each thermal has an associated 3-dimensional velocity and position, which specifies the potential energy/altitude gain available to the glider (defined as the effective maximum height) and the thermal radius of effect (estimated as one-tenth of the effective maximum height). The glider parameters of interest are outlined in Table 1.

Table 1: Glider platform and sensor parameters

Parameter	Value
Nominal gliding speed	25 m/s
Descent rate	1 m/s
Maximum banking angle	55°
Sensor range	1000m
Sensor sample interval	60s
Camera FOV	20° – 30°

The sensor range of 1000m is an assumption based on the work in [8] which presents a method for state estimation of non-linear wind fields. The time step used in [8] was 1.2 hours and dealt with lower frequency dynamics of the wind field, however for our purposes, we are concerned with the high frequency dynamics and thus have chosen a 40 second horizon for the sensor range. The reader is referred to [6] for further simulation results that show the effect of varying sensor ranges and platform gliding dynamics.

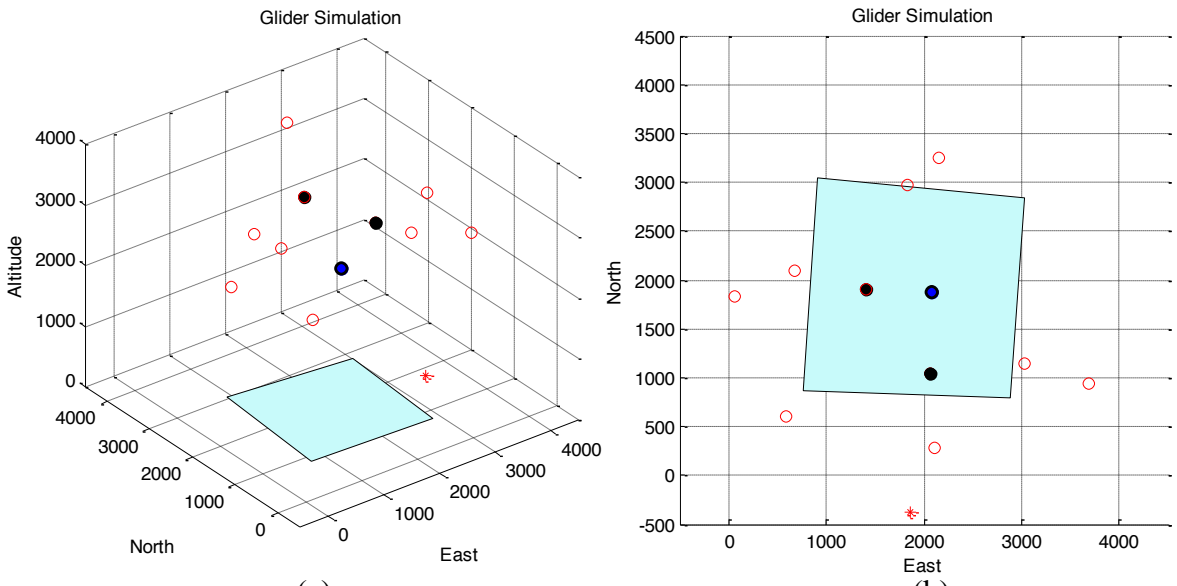
The waypoints for the glider mission are predefined and loaded at the start of the simulation. Two main missions were examined in simulation:

1. Local area survey using a boustrophedon search pattern.
2. Cross-country exploration following a predominantly straight path.

These two paths best represent the ARES scout mission objectives of sampling regional atmospheric and crustal properties, and following water trails to test for hydrogen content as outlined in the introduction.

Fig. 4 gives a snapshot of the Matlab simulation. In the instance shown, the glider is adjusting its course to intercept the thermal at (1020, 2024, -3400)[1].

[1]



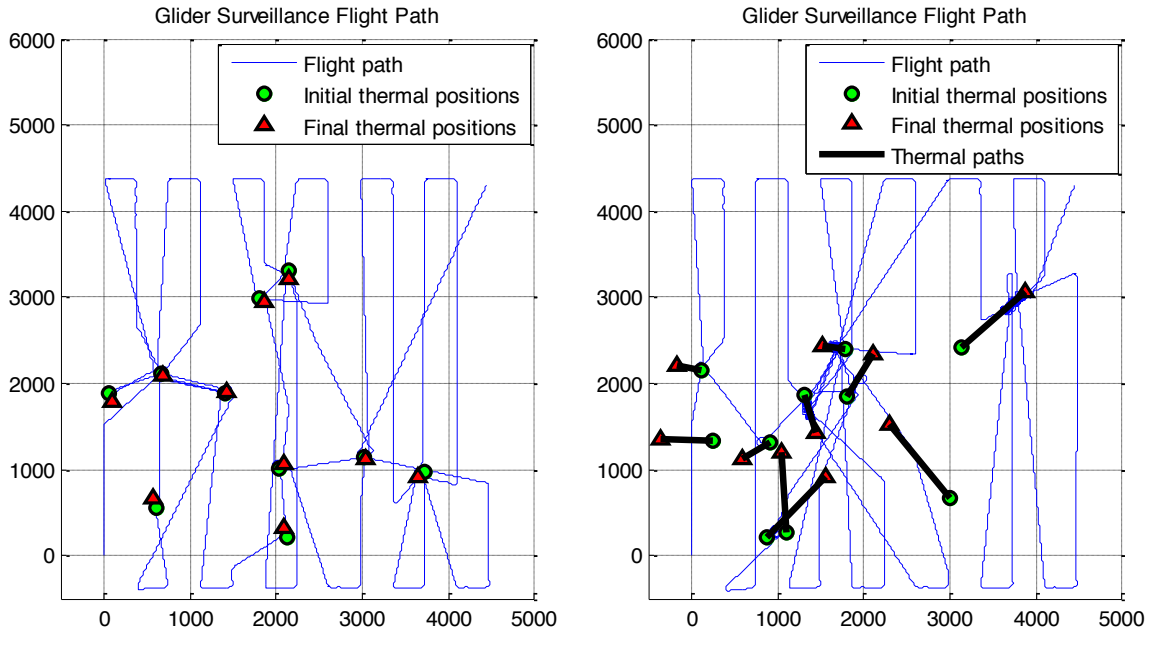
(a) (b)

Fig. 4: Matlab simulation of the glider mission; (a) 3D view, (b) aerial view. The glider is represented as the blue point mass with a camera footprint shown in light-blue; red circles represent effective thermal heights and are located at the centre of the thermal columns; black points mark assumed thermal locations; red stars represent the flight plan waypoint.

Simulation Results

Localised Survey

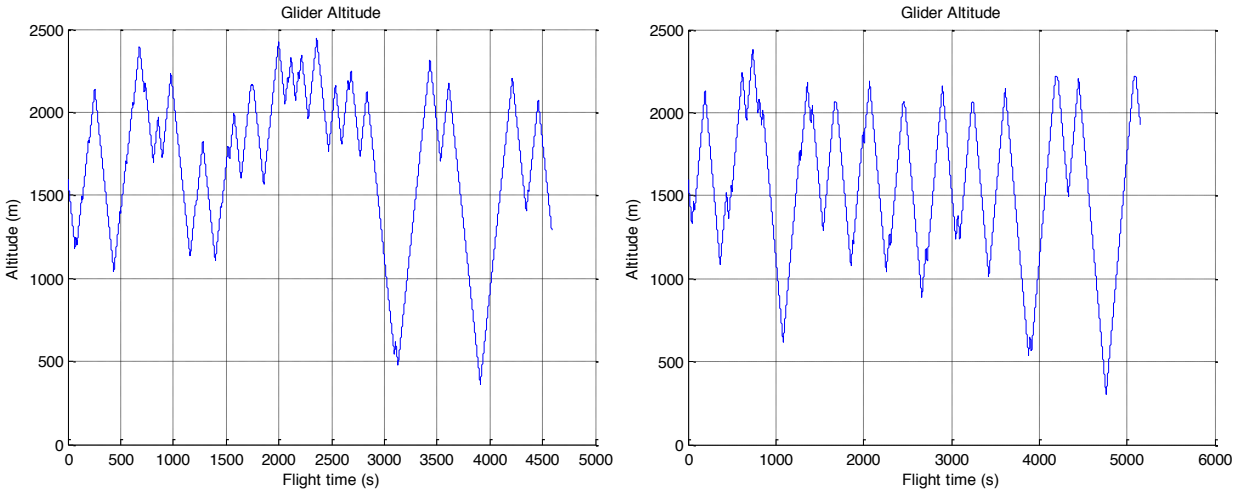
The flight paths taken by the glider under the localised survey mission mode are shown in Fig. 5. In both flights, the BFS decision model selects intuitive thermal waypoints to maintain altitude and navigate the surveyed region. It is evident from Fig. 5b that the system is capable of predicting the movement of the thermals as previously visited thermals travel significant distances (up to 500m) between survey loops.



(a) (b)

Fig. 5: Flight path taken by the glider to survey a region of slow-moving thermals (a)*; flight path taken by the glider to survey a region of fast-moving thermals* (b).

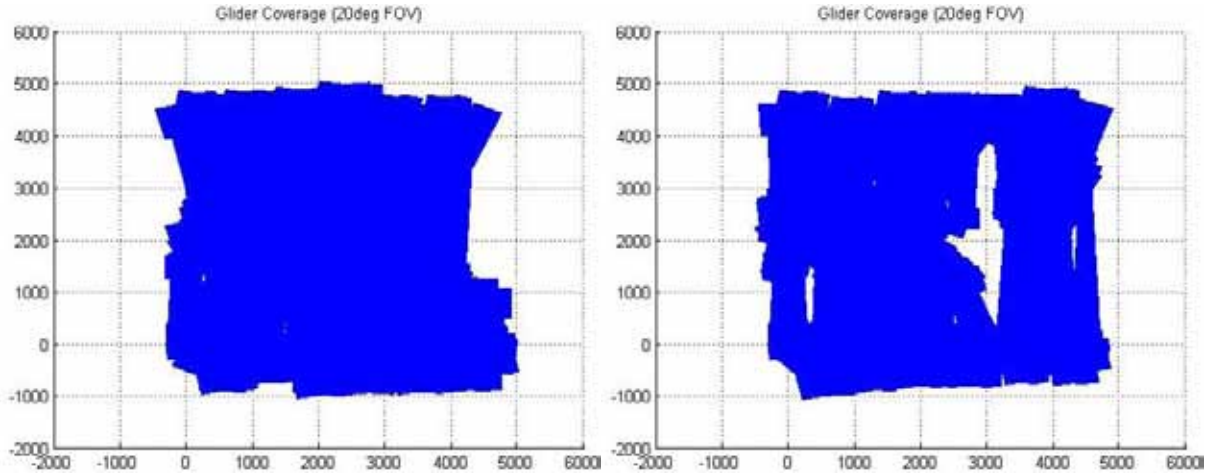
The flight paths show that the basic boustrophedon search pattern is maintained while the glider exploits thermal sources for altitude gain. Fig. 6 shows the glider altitude over the flight. For both flights, the glider remains near the nominal gliding altitude of 1.5km, thus maintaining image resolution whilst correcting when the critical flight altitude of 750m was violated.



(a) (b)

Fig. 6: Glider altitude over the flight paths shown in Fig. 5a and Fig. 5b, respectively.

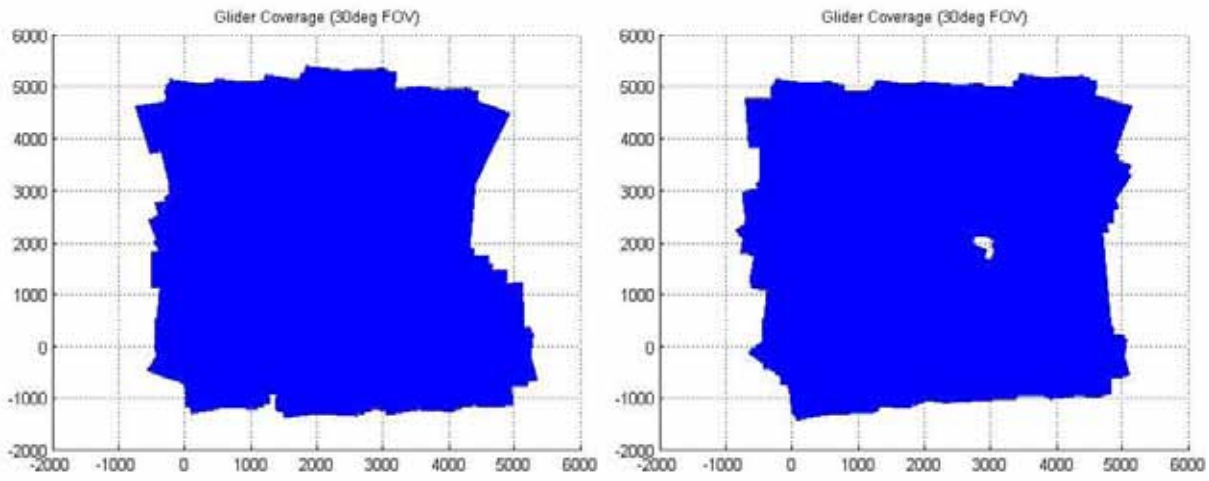
The 20° FOV sensor coverage over the flight path is given in Fig. 7. Fig. 7a shows that apart from a few isolated pockets, the entire 4000m×4000m survey region is captured. In Fig. 7b, the gaps in the sensor coverage are due to the placement of thermals within the field causing the glider to deviate further from the nominal search path to gain energy from nearby thermals. Despite these path deviations, over 90% of the search area was captured.



(a) (b)

Fig. 7: 20° FOV sensor footprint coverage across the flight paths shown in Fig. 5a and Fig. 5b respectively. Observations taken at a glider bank angle > 10° are discarded due to poor image resolution. Such images would be taken when the glider is turning around waypoints or circling thermals and would be subject to higher levels of noise.

Given the 30° FOV sensor, the coverage across the flight paths is shown in Fig 8. The entire surveyed region is now captured in case (a) and Fig. 8b shows a significant improvement from the 20° FOV camera in case (b). Over 99% of the surveyed region is captured with the wider lens camera. It is also noted the poorer quality sensor observations taken while the glider is circling may still be used to resolve coverage 'black spots' if necessary.

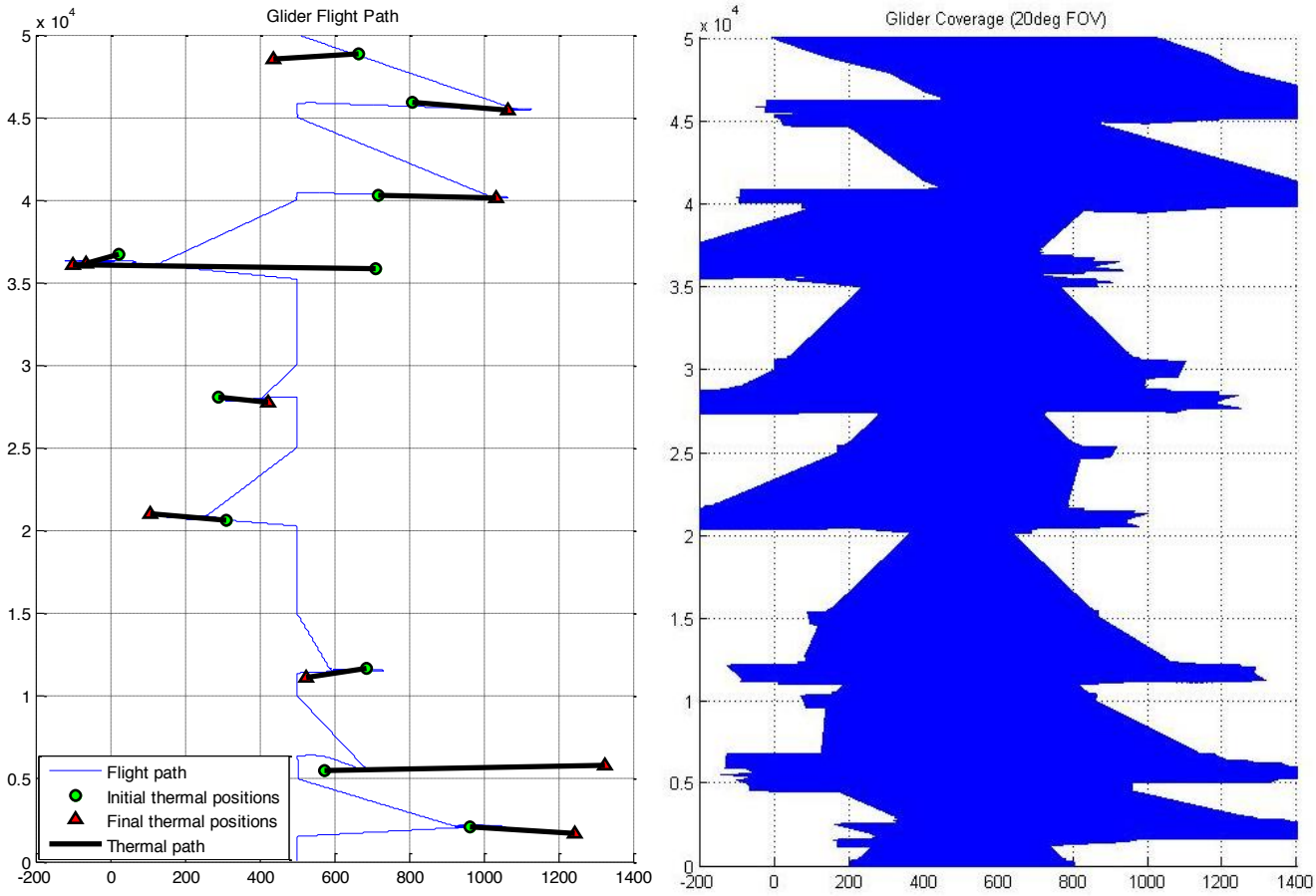


(a) (b)

Fig. 8: 30° FOV visual camera footprint coverage across the flight paths shown in Fig. 5a and Fig. 5b respectively.

Cross-Country Exploration

The performance of the system while the glider is operating under the cross-country exploration mode is shown in Fig. 9. Fig. 9a shows the glider's ability to maintain its search track whilst locating and exploiting nearby thermals to increase gliding endurance. Fig. 9b shows that the 20° FOV sensor footprint provides consistent coverage along the nominal path.



(a) (b)

Fig. 9: Glider performing autonomous exploration traversal over a 50km straight stretch.

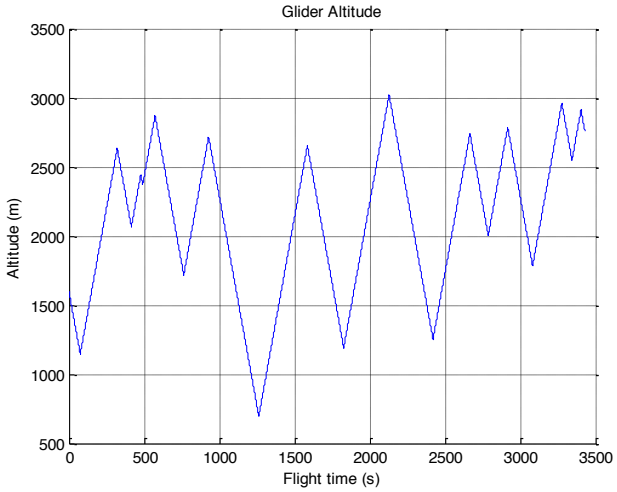


Fig. 10: Glider altitude over the cross-country flight path shown in Fig. 9a.

The glider altitude over the 50km flight is shown in Fig. 10. The maximum acceptable glider altitude was increased to 2500m for this simulation to ensure glider safety; due to this, the glider's altitude varies mainly between 1000m to 2500m over this flight plan. It is noted that over the course of the flight, the glider only drops below the critical altitude once and rarely deviates further than 500m off the allocated track.

Real-Time Computation

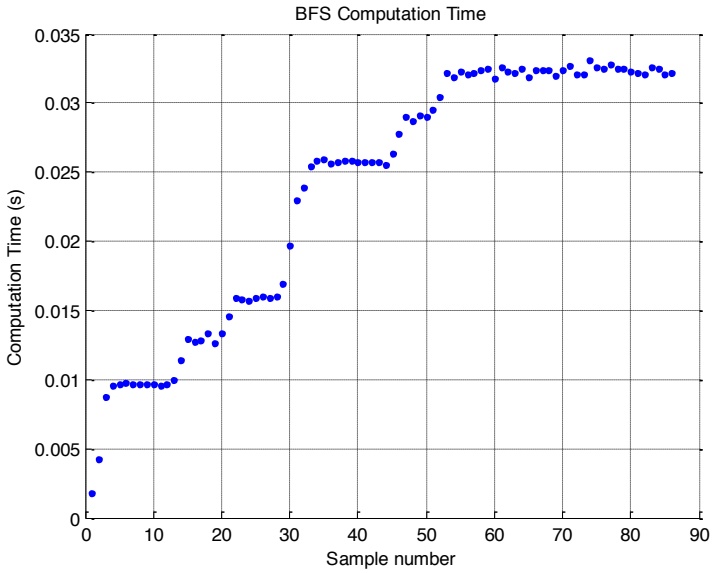


Fig. 11: BFS computation time over the flight plan shown in Fig. 5a. Simulation was run on an Intel® Core™ 2 Duo CPU E8400 @ 3.00GHz 4.00GB RAM.

Fig. 11 shows that the BFS computation time scales proportionally to the number of thermals detected by the glider. It is evident from the graph that although previous sensor observations and forecasts are retained in the BFS, the computational load does not increase over samples when no new thermals are detected. The BFS computation time of 3.25ms/thermal is compatible to the sensor sample rate (once every 60 seconds) and the dynamics of the glider. However, over extended flights, it will be desirable to remove thermals that have not been observed for a period of time to free on board memory space and reduce the computational load. These thermals may have moved out of the region of interest or have dissipated their energy altogether.

Conclusion

The application of a Bayesian forecast-decision system to an autonomous glider operating in an unknown and dynamic environment has been shown in this paper. The BFS is capable of seeking out and exploiting available thermal energy sources in the environment to maintain gliding altitude whilst considering an external mission goal such as the ARES glider missions of regional surveys and cross-country exploration.

Future work in this area may involve the use of terrain modelling from camera observations to enable another layer of probability modelling focusing on the coupling between geographical features such as mountain ranges and ridges and sources of rising air (orographic effect).

References

1. Levine, J.S., et al. 2003, "Science from a Mars Airplane: The Aerial Regional-scale Environmental Survey (ARES) of Mars", 2nd AIAA "Unmanned Unlimited" Systems, Technologies, and Operations – Aerospace, Land, and Sea Conference, Paper Number AIAA-2003-6576.

2. Guynn, M.D., et al. 2003, "Evolution of a Mars Airplane Concept for the ARES Mars Scout Mission", 2nd AIAA "Unmanned Unlimited" Systems, Technologies, and Operations – Aerospace, Land, and Sea Conference, Paper Number AIAA-2003-6578.

3. Sandford, S.P., et al. 2003, "ARES and Beyond: Autonomous Aerial Platforms Provide a Unique Measurement Capability for Earth and Planetary Science", 2nd AIAA "Unmanned Unlimited" Systems, Technologies, and Operations – Aerospace, Land, and Sea Conference, Paper Number AIAA-2003-6579.

4. Levine, J.S. 2010, *ARES Mars Scout Mission Proposal – Science*, Hampton, viewed 27 June 2010, < http://marsairplane.larc.nasa.gov/science.html>.

5. Krzysztofowicz, R. 1985, "Bayesian Models of Forecasted Time Series", *Water Resources Bulletin*, vol. 21, no. 5, pp. 805-814.

6. Chung, J.J. 2009, "High Level Risk Analysis and Decision Making for an Autonomous Glider", BE Hons thesis, University of Sydney.

7. Raiffa, H. & Schlaifer, R. 1961, Applied Statistical Decision Theory, 3rd edition, Harvard University, Boston.

8. Park, S., et al. 2009, "Learning Covariance Dynamics for Path Planning of UAV Sensors

Field Strength Measurement of VLF Radio Wave Propagation at 19.8 kHz between Australia and India

*C. T. More [1], A. K. Sharma [2] and R. V. Bhonsle [2]
Kenneth J.W. Lynn[3]

[1] *Department of Physics, Miraj Mahavidyalaya Miraj, Dist-Sangli, 416410, India*
[2] *Department of Physics, Shivaji University, Kolhapur, 416 004, India*
[3] *Ionospheric Systems Research, 16 Heritage Dr., Noosaville 4566, Australia*
**E-Mail for correspondence: chandrakanttmore@yahoo.com*

Summary: The purpose of this experiment was to study the effect of a solar activity on the field strength of VLF radio wave at 19.8 kHz reflected from D-layer of the Ionosphere. For this, hexagonal loop antenna of 1.5-meter diameter is connected to the VLF receiver tuned to 19.8 kHz. The VLF field strength monitoring system is installed at Khatav, India (16°46' N, 75°53' E). The monitor was used to receive 19.8 kHz transmission by VLF radio station NWC at North West Cape, Western Australia (22°49' S, 114°25' E). The field strength of the VLF radio wave transmitted by the above station via ionosphere was continuously monitored for more than one year at 5 seconds interval. The continuous monitoring of these VLF radio waves from July 30, 2009 to August 31, 2010 clearly shows diurnal and seasonal variation in the field strength. It also shows the effect of solar X-ray flares and annular solar eclipse of January 15, 2010 on the field strength of VLF radio wave propagation. In this paper the analysis and interpretation of the data received is discussed.

Key words: VLF radio propagation, diurnal and seasonal variation, solar X-ray flares, annular solar eclipse.

Introduction

The study of the effects of solar activities like solar flares, sunspots, solar wind particles, coronal mass ejections on the ionosphere of the Earth is very important as they all affect the Earth and its environment. The activities such as UV and X-ray radiation affect VLF (3-30 kHz) radio wave propagation, submarine communication and deep sea navigation. The VLF spectrum has many interesting monitoring possibilities due to its unique characteristics. Field strength monitoring of VLF radio wave transmission via ionosphere of the Earth is an important ground based tool to study solar X-ray flares and also their effects on ionospheric VLF radio wave propagation [1].The theory of VLF waveguide propagation is at a stage of maturity and thus can be used to interpret a large number of different geophysical phenomena observed on VLF paths [1, 2, 3].

The ionosphere of the Earth is formed due to solar radiation like X-rays, EUV rays as well as cosmic X-radiation. The several ionospheric layers such as D, E and F layers are created with increasing electron densities at different heights [1, 2, 3]. They are also called regions. The D-layer (70-90 kilometers) which is ionized by X-rays of 0.1 to 1 nm, E-layer (100-120 kilometers) ionized by EUV (80-103 nm and X-rays of 1-20 nm) and F-layer forms F1 and

F2 layers during the day ionized by EUV of 20-80 nm (above 120 kilometers) [4]. Each layer has its own electron density profile with height and their structure changes daily and seasonally under the influence of the sun. The ionospheric behaviors play a role in radio propagation that varies strongly with the frequency. Thus the propagation of VLF radio communications is affected by the ionospheric disturbances At VLF both ground and ionosphere are good electrical conductors and form a spherical earth-ionosphere waveguide and hence the VLF waves transmitted by radio stations propagate by waveguide mode [4, 5, 6]. The VLF radio wave field strength monitoring system (SID Monitor) installed at Khatav (India) captures the 19.8 kHz waves transmitted by VLF radio station NWC Cape North Australia via D-layer in daytime and via E- and/or F-layer in nighttime at the reflection points in the ionosphere.

Recording and monitoring of VLF signal

The observations reported here were recorded by the VLF field strength monitoring system which is developed by Stanford Solar Center, Stanford University, USA [7] and is situated at Khatav, India ($16^{o}46'$ N, $75^{o}53'$ E). The receiver has been in stable operation since July 2007, providing well calibrated data from VLF radio station NWC at North West Cape, Western Australia ($21^{o}46'$ S, $114^{o}44'$ E). With the continuous operation, the received signals reveal regular, diurnal and seasonal field strength variations under quiet and disturbed ionospheric conditions caused by solar activity. Enhancements of signal strength due to increased reflectivity of D-layer caused by even minor solar X-ray flares are very well recorded with this SID monitor at 19.8 kHz.

The Experimental set up

Fig. 1 shows block diagram of VLF field strength monitoring system which consist of a loop antenna, SID monitor, A/D Converter and Computer.

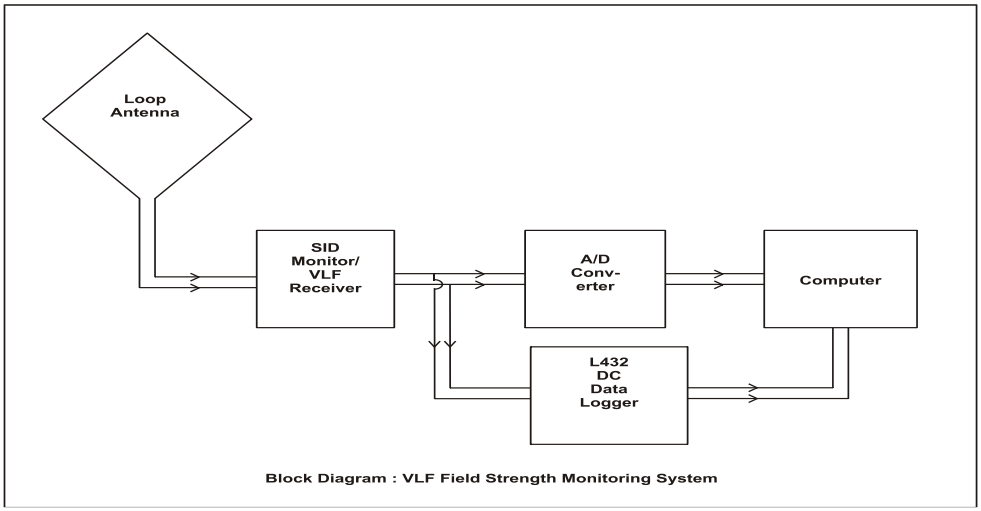


Fig. 1: the block diagram of 19.8 KHz VLF Field Strength Monitoring System

The antenna here used is a hexagonal loop antenna of 1.5 meter diameter, which is made up of 25 turns of 19 gauges super enameled copper wire. The antenna is connected to the receiver by RX-59 coaxial cable and placed at a distance of 45 feet from the receiver. The VLF field strength monitoring system and loop antenna is shown in figure 2. The SID monitor is nothing but VLF amplitude modulation (AM) receiver and it is tuned to NWC Cape North VLF radio station, Australia which is operating at 19.8 kHz. The signal captured by loop antenna is detected, amplified, rectified and the receiver output measures signal strength in DC voltage. The signal is scaled to swing between the two power supply voltages, thus giving the meter as overall 10 volt range (+ or – 5 volt) [9].

(a) *(b)*

Fig. 2: 19.8 kHz VLF Field Strength Monitoring System (a) & Hexagonal Loop Antenna (b)

The signal strength data is stored automatically in a computer through A/D converter by using software at 5 seconds interval round the clock.

Calibration of SID Monitor

Fig.3 shows an experimental arrangement to determine the response of the SID Monitor to the input signal of 19.8 kHz obtained from the function generator. For this purpose, loop antenna was replaced by the function generator. For small input signal, the receiver obeys square law but beyond a certain level of input signal, response of the receiver obeys linear law. The SID Monitor gain or amplification factor is given by the slope of the response curve and is equal to 800 for the input voltages smaller than 4mV. Due to this high gain, the SID Monitor can also receive weak signal.

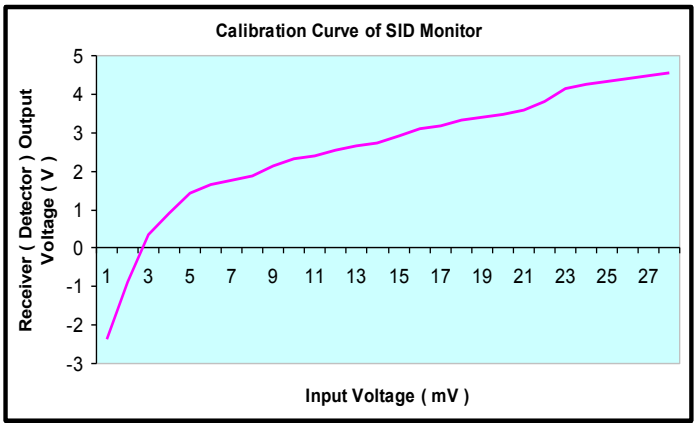


Fig.3: Calibration curve of 19.8 kHz SID Monitor

By using this experimental setup, many C-class and M-class solar X-ray flares are detected in the form of enhancement of VLF field strength. The data recorded shows daily sunrise, sunset effect very clearly and variation of field strength of each day round the clock. The loop antenna is directed towards south east direction so that antenna picks up maximum signal is received.

Diurnal and seasonal variation

Diurnal variation

The distance between the NWC 19.8 kHz transmitting station in Australia and Khatav (India) is around 6184 km. Since single-hop VLF radio wave reflection via D-layer does not reach beyond 2000 km [3] an earth-ionosphere waveguide mode of propagation is assumed [4, 5, 6]. The frequency is such that only the first 2 waveguide modes need be considered at night with the first mode probably predominant. Any direct effect of the second mode on this path has yet to be determined.

The word "diurnal" simply means "throughout the day. The diurnal variation of the VLF reflection height is very simple and repetitive. During the day, the reflection height depends almost exclusively on the zenith angle of the sun [3]. This dependence can be seen in fig.4 as a smooth change of field strength in daytime with a maximum at mid-day. At night when the D region has disappeared, the reflection occurs from the very weak night-time E-layer at a height of about 90km. The E region base of the ionosphere at night is less stable than in daytime.

Diurnal variations of VLF signals propagated over long distances (>5 Mm) were first observed by *Yokoyama and Tanimura* [9] in 1933 with diurnal phase variations first reported by Pierce in 1955 and Crombie et al in 1958. During daytime, solar UV and X-rays ionize the neutral atmosphere creating the D, E and F layers. The D layer is created due to ionization of NO by solar L-α (1216Å) radiation. The variation in field strength of 19.8 kHz radio wave on July 30 to 31, 2009 at Khatav, India is shown in fig. 4.

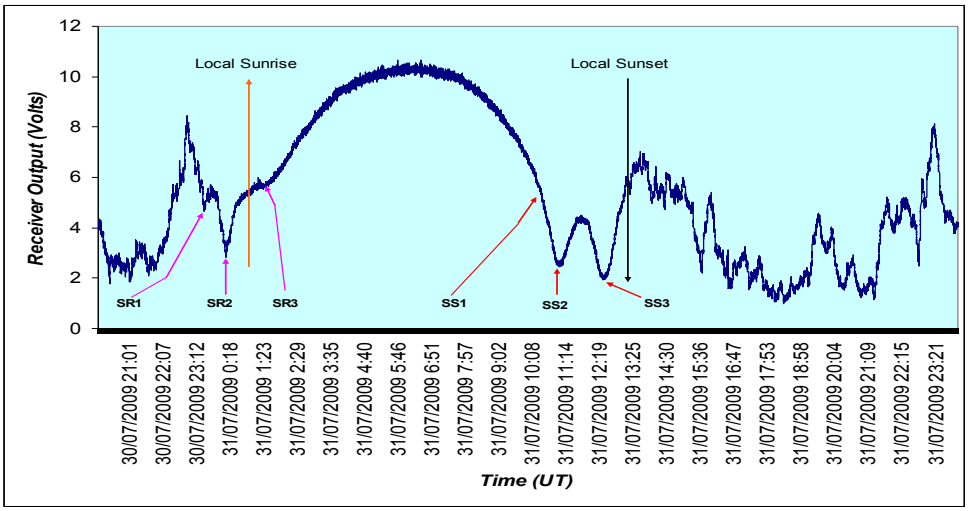


Fig. 4: Typical diurnal variation of NWC signal strength (19.8 kHz) received at Khatav (India) clearly showing sunrise and sunset on July 30 to 31, 2009

The distance between the transmitter (NWC Cape North Australia) and receiver (Khatav, India) is 6184 km. In Fig. 4, the first signal minimum (SR_1) observed at the receiver, as the sunrise terminator moves westward over the propagation path, is observed at 23:23 UT (30, July 2009) with a subsequent signal minimum at 00:03 UT (SR_2). Following the completion of sunrise over the whole path, an increase in signal strength occurs as a temporary C layer develops [3, 8] when sunlight releases electrons from negative ions built up during the night. A slight kink in signal strength is present (SR_3) at 03:04 UT (July 31, 2009) as the D layer begins to build and ultimately swamps the fading C layer, taking over the dominant reflection role. Maximum signal strength occurs in the middle of day when the sun reaches its highest solar zenith angle. Ionizing solar UV on the path is at its maximum at this time and the D region consequently reaches its highest electron density.

Just as two signal minima SR_1 and SR_2 occurred during the path sunrise transition, two signal minima, SS_2 and SS_3 in Fig.4 are observed at 10:19 UT and 10:54 UT (July 31, 2009) during the path sunset transition. The signal minima are produced by modal interference generated at the sunrise and sunset height discontinuities in reflection height as they move along the path [3, 10]. After sunset, the reflection of VLF signals occurs from the E region at a height of around 90 km. This very weak night-time E region is sufficient to reflect VLF radio waves whereas HF radio waves passed through to reflect from the much higher night-time F2 layer. The VLF signal strength is highly variable throughout the night due to the instability of the night-time E layer. The times of local sunrise and sunset are also shown in Fig.4 at 00:28 UT and 13:29 UT respectively.

Somewhat similar “rapid changes of amplitude occur near sunrise and sunset along the path” were observed on a Hawaii-Boulder path (5374 km) on 19.8 kHz [1] with further examples in [10, 3]. The diurnal variation in the signal strength typically repeats day after day.

Seasonal variation

Fig. 5 shows seasonal variation in the field strength of the VLF radio station NWC observed at Khatav (India) for the months from August 2009 to August 2010. The recordings were not done in the months of October 2009 and May 2010 (As the receiving system was under repair in the month of October, 2009 and the transmission from VLF station NWC was not available in the month of May, 2010)

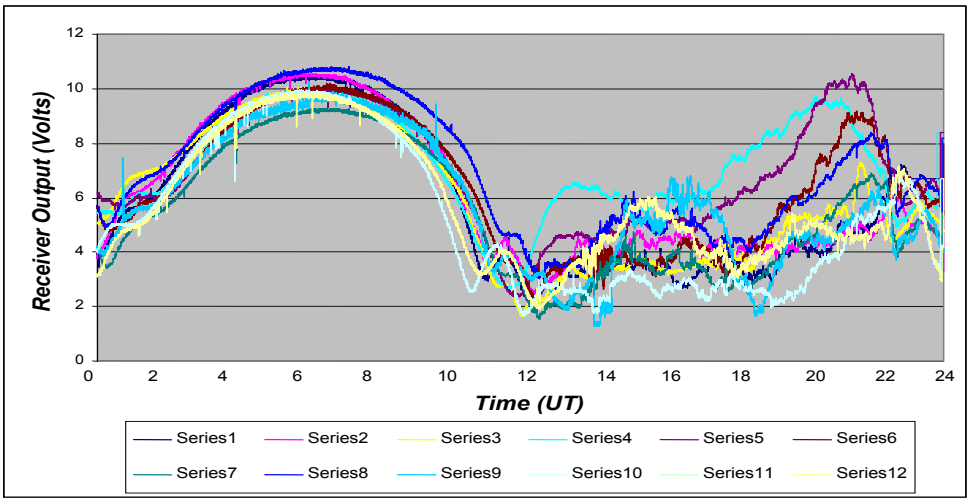


Fig. 5: Monthly average curves of undisturbed days during the period August 1, 2009 to August

30, 2010 excluding the month of October 2009 and May 2010

As explained earlier one can obtain field strength variation at the receiver input by dividing detector output voltage by the amplification factor of the receiver. The monthly average curves of undisturbed days are superimposed in one diagram for easy comparison to bring out seasonal variation.
In table 1 maximum daytime values of detector (receiver) output voltage and field strength at the input of SID Monitor are plotted.

Table 1- Seasonal variation of maximum daytime field strength of 19.8 kHz transmission (during the period from August 2009 to July 2010) recorded at Khatav (India)

Month/Year	Maximum daytime detector output(Volts) V	Field strength at the input of SID Monitor mV
August 2009	10.46	1.31
September 2009	10.43	1.32
November 2009	09.92	1.24
December 2009	08.28	1.13
January 2010	08.17	1.12
February 2010	08.76	1.19
March 2010	09.20	1.15
April 2010	09.67	1.21
June 2010	09.82	1.23
July 2010	09.93	1.24

From table 1 it is observed that during the month of August 2009, the signal strength is a maximum and during the month of January 2010, it is a minimum.

Effect of solar X-ray flares

Solar flares which have a strong hard UV and x-rays component have long been known to increase ionization in the D region resulting in a further increase in VLF phase velocity as the reflection height lowers [1, 2, 3, 7, 9]. The VLF signal of 19.8 kHz from NWC is continuously being recorded by a SID monitoring system [7] since April 2007 and shows characteristic variations in diurnal and seasonal field strength. It has been observed that solar flares are huge explosions in the Sun's atmosphere. They appear to instruments as bright flashes in visible light, often followed by a burst of high-energy protons and radiation. Moreover their characteristics can include burst of radio waves, EUV and X-rays [1, 2, 3]. A large solar flare can release a thousand million megatons of energy (more precisely 10^{28} to 10^{34} ergs) in a single explosion. The released energy is transformed into: 1) thermal energy leading to an increased brightness of e.g. the Hα and X-ray emission), 2) particle kinetic energy leading to the acceleration of electrons to energies of 10 keV to 1 GeV and ions to energies from a few MeV/nuc toGeV/nuc,3) mechanical energy leading to several kinds of plasma ejecta. Solar flares sometimes occur together with other signatures of solar activities e.g. prominence eruptions, CME's and interplanetary shock waves. However the exact relationship between these phenomena is not yet completely understood [1]. The most prominent change in the earth-ionosphere waveguide is the diurnal and seasonal variation i.e.

day-to-night (and reverse) change [1]. However, significant modifications of the propagating conditions happen due to severe changes in the lower ionosphere electron density, induced by solar X-ray flares [1, 2, 3]. Although the main source of ionization the Lyman-α emission (1215.67 Å), is enhanced during the flare event, the X-ray emission overwhelms its effect several times, leading to the increase in the D-layer electron density by 1–2 orders of magnitude. Enhanced D-layer density causes the change in the electrical conductivity at the upper waveguide edge along the trace of the VLF signal and consequently, gives rise to the change in all propagating parameters. These changes are clearly detected in the form of VLF field strength variation. Although flare effects on the VLF signal are always well recognizable, they can vary substantially along different signal traces. The theory of VLF propagation through the earth-ionosphere waveguide, at regular (quiet) ionospheric conditions, is well established [5, 6]. When solar flares occur, sudden energy bursts in the X-ray domain appear most distinctly impressed on the VLF signal, by amplitude with an abrupt increase, followed by the subsequent signal recovery within time intervals (typically less than an hour), which correspond to flare duration [8]. Over 50 flare events were detected during the period of July 1, 2009 to July 30, 2010. We relate the VLF signal variations measured by the VLF field strength monitoring system to the solar X-ray irradiance, as monitored by the GOES 12 satellite. Fig. 6 gives representative example of flare-induced field strength variations, measured for the NWC signal, on the active day of December 22, 2009.This sudden field strength variation is related to the C-5 Class solar x–ray flare detected by GOES X-ray satellite which is shown in fig 8 by the mark 1 in blue colour.

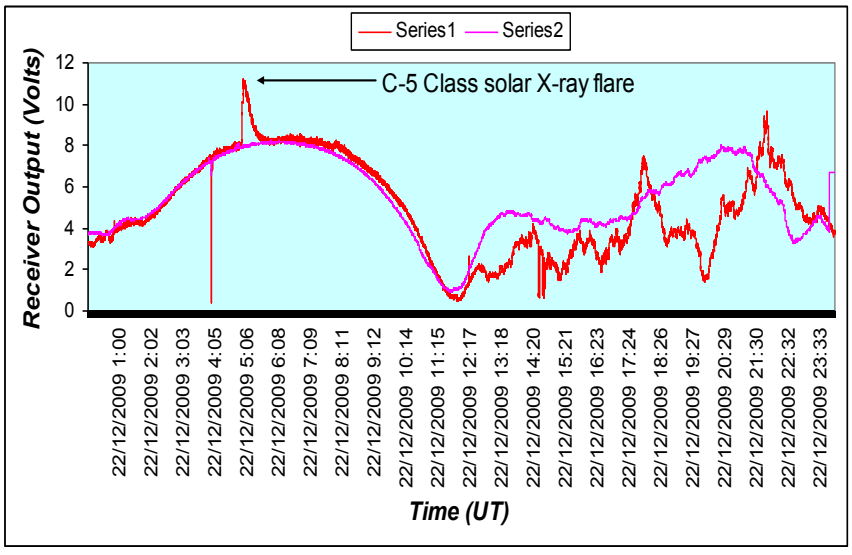


Figure 6: Diurnal field strength variation of NWC Cape North Station. Series 1- red colour curve corresponds to C-5 Class Solar X-ray Flares occurred on December 22, 2009.Series 2- pink colour curve shows the average diurnal field strength variation for the month of December 2009.

Fig. 6 shows the unperturbed average plot of field strength of the month of December 2009 is overplotted with the plot of disturbed day December 22, 2009. This perturbation is observed due to C-class solar X-ray flare. The correspondence between the X-ray flare events detected by GOES 12 satellite enhancement is remarkable; moreover, the field strength response to the flare is clearly present. Fig. 7 gives another example of flare-induced field strength variations, measured for the NWC signal, on the active day of December 23, 2009.This sudden field

strength variation is related to the C-5 Class solar x–ray flare detected by GOES X-ray satellite which is shown in fig 8 by the mark 2 in blue colour.

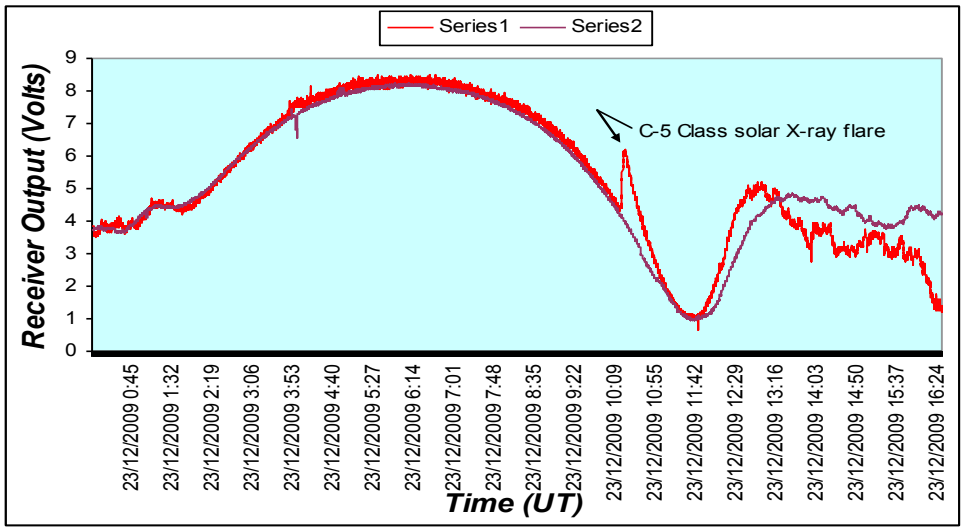


Figure 7: C-5 Class Solar X-ray Flares occurred on December 23, 2009

The Fig. 8 shows the solar X-ray irradiance as monitored by GOES 12, from December 22 to 24, 2009 (UT), indicating a sequence of flares of class B and class C. The features of enhanced
X-ray irradiance have a distinct impact upon VLF field strength characteristic, which display different patterns peculiar to the NWC path, throughout a single active day.

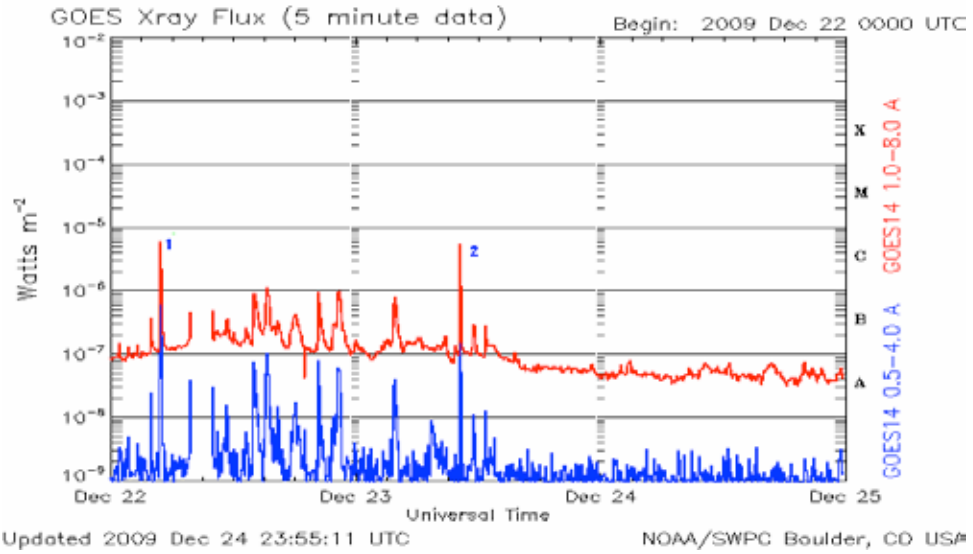


Fig. 8: GOES X-ray flux plot on December 22 to 24, 2009 Figure 9: Optical photo of sun on Dec. 23, 2009

During a solar flare, VLF waves travel at a reduced incidence angle to the earth and ionosphere as the D region expands downward thus narrowing the waveguide. The signal

strength usually increases because the wave does not lose energy as it reflects from bottom of the D-layer. However, the VLF signal strength during flare can sometimes decrease. As soon as X-ray flares end, the sudden ionospheric disturbance (SID) starts to decay and the electron density in the D-layer returns to normal as a result of attachment and /or recombination processes. The VLF field strength monitoring system has also recorded many B-class solar X-ray flare events along with those of M and X-class. The plot of signal strength against time recorded on December 22, 2009 can be compared with the GOES satellite data graphs available at http://www.swpc.noaa.gov/ftpdir/plots/x-ray. We can also confirm the H-alpha line images of the sun on the same day and correlate the event and X- ray flare. We can also do correlation study between the sunspots and solar X-ray flares.

Multiflare analysis

For the analysis of VLF data, a series of flares were selected that occurred on January 20, 2010 and they are presented, along with the average VLF field strength measurements, in figure 9. The reasons for this choice are as follows: 1) the flare intensity (ranging from class C to M) is characteristic for the majority of the flares that have been VLF recorded at Khatav (India) on this active day. (2) many SIDs in the form of signal enhancements overlapping in occurrence due to multiple solar X-ray flares have been recorded on this day from sunrise to sunset.

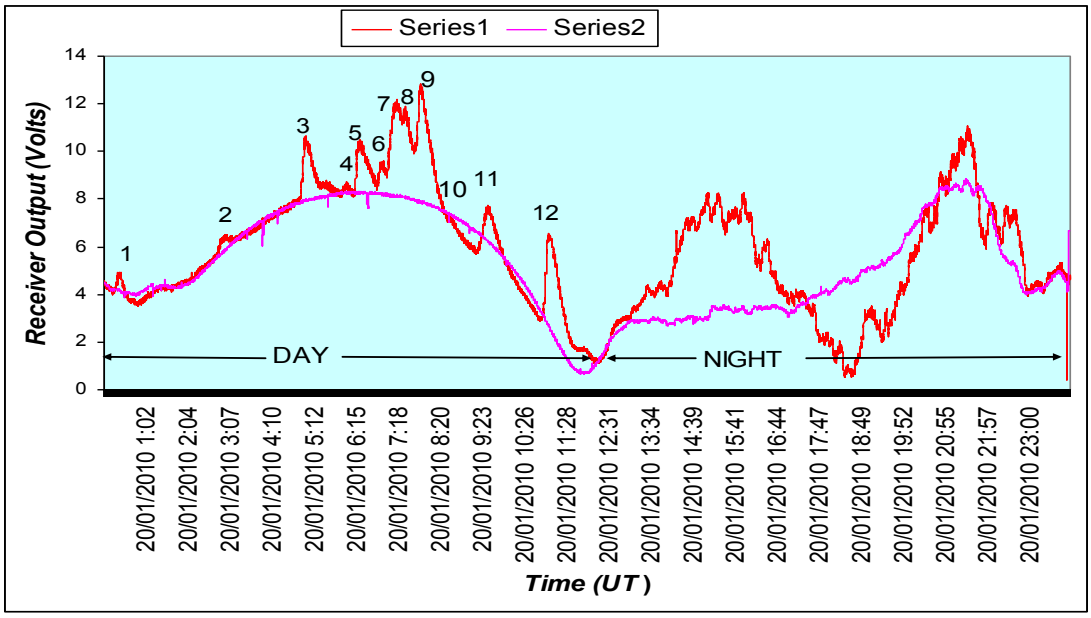


Fig. 9: Multiflare occurrence on active day of January 20, 2010

Fig. 10 shows the GOES 12 satellite X-ray plot on active day of January 19, and 20, 2010
It is observed that solar X-ray flare events started to occur from January 18, 2010 and these solar activity ends on January 21, 2010. It is also observed that many M-class and C-class solar X-ray flares have been occurred during this period.

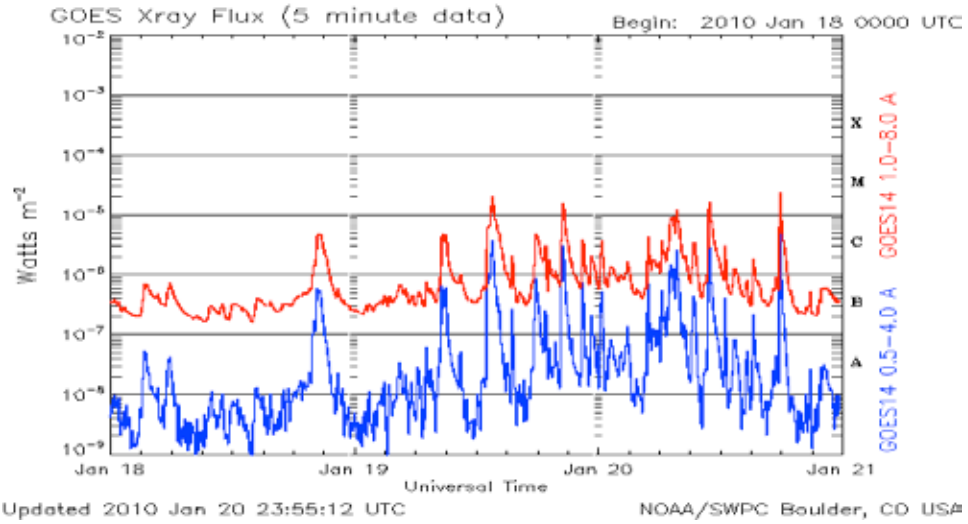


Fig.10: GOES 12 satellite X-ray plot on active day of January 19, and 20, 2010

The characteristics of the flare events are given in Table 2: The first column identifies the flare number the second column lists the time at which the characteristic feature of the field strength variation (minimum/maximum) is detected. The third column shows the duration of events.

Next column four shows the output voltage of detector when flare occurred and column five shows the average detector output voltage for the month of January 2010. The column six shows variation ∂V in detector output. The column seven shows the Class of flare and the last column shows the Flare Intensity (W/m^2) as the data recorded by satellite GOES 12.

Table 2: Measured field strength variation at different stages of flare occurrence and at different angle of elevation and intensity of X-ray flares

Solar Elevation Angle	Flare No.	Flare Time (UT)			Duration (Minutes)	V_f (Volts)	Vq (Volts)	∂V (Volts	Class of flare	Flare Intensity (W/m^2)as satellite GOES 12 data
		Start	Max	End						
21°	1	00.13	00.23	00.30	17	4.82	4.11	0.72	C4	4×10^{-6}
58°	2	02.49	02.58	03.13	24	6.44	3.81	2.63	C1	1×10^{-6}
73°	3	04.50	04.58	05.22	32	10.47	8.01	2.46	C1	1×10^{-6}
77°	4	05.51	06.00	06.11	20	8.54	8.27	0.26	C6	6×10^{-6}
72°	5	06.13	06.20	06.46	33	10.25	8.24	2.02	C5	5×10^{-6}
64°	6	06.46	06.53	07.00	14	9.57	8.21	1.36	C7	7×10^{-6}
61°	7	07.00	07.16	07.23	23	12.17	8.14	4.03	C8	8×10^{-6}
56°	8	07.23	07.28	07.41	18	11.72	8.05	3.67	C9	9×10^{-6}
52°	9	07.41	07.51	08.33	52	12.62	7.79	4.83	M1	1×10^{-5}
39°	10	08.33	08.39	08.46	13	7.31	7.09	0.22	C2	2×10^{-6}
30°	11	09.15	09.30	10.01	46	7.55	5.82	1.73	C5	5×10^{-6}
8°	12	10.50	11.03	11.38	48	6.38	1.94	4.45	M2	2×10^{-5}

Determination of field strength variation ∂V due to X-ray flares

The field strength variation ∂V is determined from the VLF field strength measurements on the quiet day V_q, and that due to X-ray flare V_f is given by

$$\{\partial V\} = \{V_f - V_q\} \quad (1)$$

It can be seen from Table 2 that the smallest enhancement in field strength (as measured by the detector output) is 0.214 volts and largest one is 4.828 volts. However, amount of field strength enhancement depends upon the intensity of X-ray flux and solar elevation angle. If the detector output voltage is divided by amplification factor of the receiver, one can get field strength variation at the receiver input terminal. Thus it is clear that our SID Monitor (VLF Field Strength Monitoring System) is very sensitive and can detect even very minor solar X-ray flares recorded by GOES 12 Satellite.

Effect of annular solar eclipse of January 15, 2010

The solar eclipse of January 15, 2010 was an annular eclipse of the Sun with a magnitude of 0.9190. This was the longest duration annular solar eclipse of the millennium with a maximum duration of 11 minutes and 7.7 seconds. It was visible as a partial eclipse in much of Africa, Eastern Europe, Middle East and Asia. It was seen as annular within a narrow stretch of 300 km width across Central Africa, Maldives, South Kerala (India), South Tamil Nadu (India), North Sri Lanka, parts of Burma and parts of China. The eclipse started at the Central African Republic, traverses Cameroon, Congo and Uganda passes through Nairobi, Kenya, and enters the Indian Ocean and reaches greatest eclipse. It was the longest on land with 10.8 minutes of viewing.

In Fig. 11, the red track shows the path of annular solar eclipse of January 15, 2010. The different ellipses and circles shown in this figure are the shadows of moon at different regions on the surface of the Earth along the eclipse path.

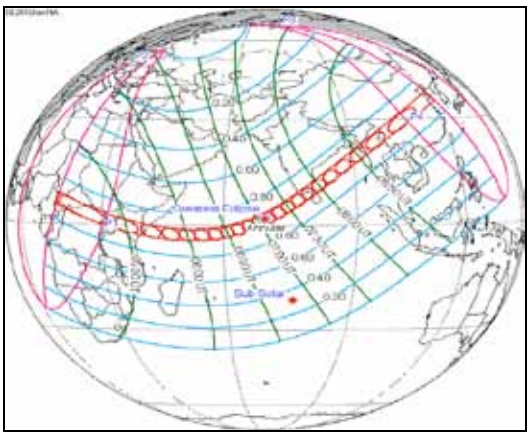


Fig. 11: Path of annular solar eclipse of January 15, 2010

On January 15, 2010, the field strength monitoring system recorded the VLF signal of NWC. The eclipse path and path of VLF waves crossed with each other over the Indian Ocean. Fig. 12 shows field strength variation on three days during the annular solar eclipse of January 15, 2010 (pink line- January 14, green line – January, 16 and red line - January 15, 2010). The

departure of eclipse day field strength from the diurnal variation of the month of January, 2010 can be seen from the figure 12 in which decrease in the field strength can be attributed to decrease of flux (Lyman α-1216 Å).

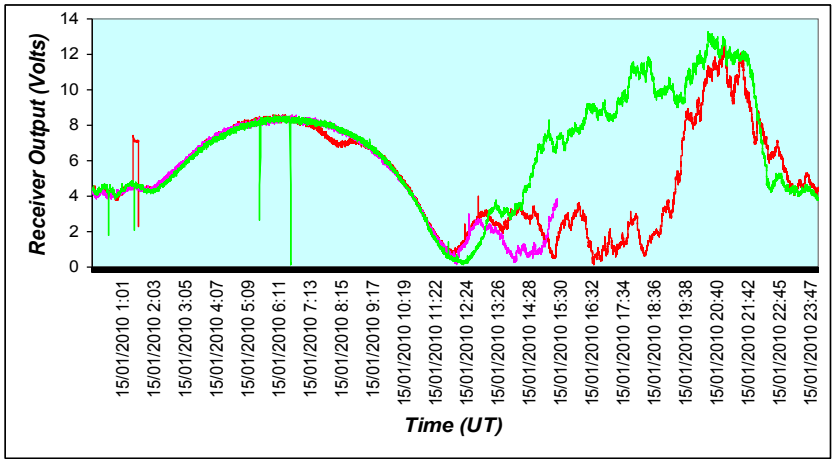


Fig. 12: Field strength variation on three days during the annular solar eclipse of January 15, 2010 (pink line- January 14, green line – January, 16 and red line - January 15, 2010)

On January 15, 2010 the field strength of NWC started to decrease its value as the moon shadow started to fall on the VLF radio wave path from NWC Australia to Khatav (India) over the Indian Ocean at 07:15 UT. The maximum dip in field strength is observed at 08.15 UT. After this dip the field strength again started to increase and it recovered its quiet day value at 09:00 UT. So the total time of effect of annular solar eclipse on VLF path from NWC to Khatav (India) is 1 hour and 15 minutes.

Conclusion

1. *Diurnal and seasonal variation*

A VLF field strength monitoring system is very useful for the study of solar activity including solar X-ray flares, sunspots, solar wind particles and coronal mass ejections. It is also useful to study the lower parts of the quiet ionosphere as observed on the field strength of VLF signals from distant VLF radio stations. The amplitude of the NWC signal as recorded at the Khatav site (India) varies only slightly from day to day during quiet days of minimum solar disturbance (from local sunrise to local sunset). The effect of mode conversion during sunrise and sunset on the field strength of 19.8 kHz NWC transmission observed at Khatav (India) is also repetitive and has yet to be studied in detail. The pattern of diurnal and seasonal variation in field strength has been described.

2. Effect of solar x-ray flares

VLF amplitude of the NWC Cape North 19.8 kHz transmitter recorded by the field strength monitoring system at Khatav (India) have been analyzed for about 50 solar X-ray flare events, during the period July 2009 to July 2010. The measured field strength variation and their characteristic features, specific for the NWC path, which is almost over Indian Ocean, have been related to the solar flare intensities, as monitored by the GOES-12 satellite. The results arrived at allow for the classification of solar X-ray flares by their effect on the lower ionosphere and indicate the possible mechanisms of VLF propagation in perturbed conditions: 1. Low C-class flares cause a small increase of field strength up to a single maximum appearing with a time delay of a few minutes after peak X-ray irradiance (flare

events C3 and C2). A likely explanation for the increase in field strength can be found in the field strength enhancement as the signal penetrates the D-layer undergoing ionization redistribution. Along with the electron density increase, the upper boundary of the waveguide drifts slightly downwards to lower heights.

3. Effect of annular solar eclipse

The variation in the field strength of the NWC transmitter was observed during the annular solar eclipse of January 15, 2010, which was attributed to decrease of UV flux (Lyman α-1216 Å) thereby decreasing reflectivity of D-layer).

Acknowledgements

The supports and encouragements from Prof. Dr.Deborah Scherer of Stanford University USA (for providing SID Monitors tuned at 19.8 and 24 KHz , data logger, software etc) as well as co-operation and encouragements from Prof. Sharad Patil , chairman of Yashawant Shikshan Sanstha Sangli (India) Prof. Dr. J. B. Chaugule (Former Principal) and Prof. Smt. Bharamgude S.S.(Principal) of Miraj Mahavidyalaya Miraj (India) are gratefully acknowledged. The authors would like to thank Prof. Dr. S. D. Khambe of Miraj Mahavidyalaya Miraj (India) for significant help in preparation of this paper.

References

1. Davies K. "*Ionospheric Radio Propagation*", NB5 Monograph 80, 1975.
2. Barr R, I., D. L. Jones, C. J. Rodger, "ELF and VLF Radio Waves", J. of Atmos. and Solar- Terr. Phys., 62, 1689 (2000).
3. Lynn, K. J. W., "VLF Waveguide Propagation: The Basics", *PROPAGATION EFFECTS OF VERY LOW FREQUENCY RADIO WAVES*: Proceedings of the 1st International Conference on Science with Very Low Frequency Radio Waves: Theory and Observations. AIP Conference Proceedings, Volume 1286, pp. 3-41, DOI: 10.1063/1.3512893, 2010.]
4. Budden K.G., **"***The Waveguide Mode Theory of Wave Propagation*'' Logos Press, 1961.
5. Wait, J. R., *Electromagnetic Waves in Stratified Media*, Pergamon, Taritown, N.Y., 1962.
6. Wait, J. R., "*Terrestrial Propagation VLF Waves, J. Res.*, NBS (National Bureau of Standards) 64D, (Radio Prop) ''1960 - 153.
7. SID Manual. ''published from the Stanford Solar Center" – 20053
8. Raulin J.P., Bertoni F.C., Gavilan H.R. Samanes J.C., "Long-term and transient forcing of the low ionosphere monitored by SAVNET", *PROPAGATION EFFECTS OF VERY LOW FREQUENCY RADIO WAVES:* Proceedings of the 1st International Conference on Science with Very Low Frequency Radio Waves: Theory and Observations. AIP Conference Proceedings, Volume 1286, pp. 3-41, DOI: 10.1063/1.3512893, 2010.]

9. Yokoyama, E., and, Tanimura, *Some long-distance transmission phenomena of Low frequency waves, Proc. IRE*, 21,263-270, 1933.

10. Crombie, D. D., Periodic fading mode of VLF signals received over long paths during sunrise and sunset, *J. Res. Natl. Bur. Stand., Sect. D, 68, 27-34*, 1964.

APPENDIX A

10ASSC
List of Presentations & Posters

List of Presentations

Abstracts are listed in alphabetical order of first author

Name of Presenter	Names of all authors	Title
Yuri Amelin	Yuri Amelin	Sequencing the early Solar System
Sudantha Balage	Sudantha Balage, Russell Boyce, Neil Mudford and Sean O'Byrne	Near-Body Asymptotic Similarity of Hypersonic Bluff Body Flows – Gas Dynamics of Planetary Entry Capsules
Rasit Baykara	Rasit Baykara, Alistair Moore, Alex Fairlie, James Nour, Jasim Ahmed, *Victor Djamovski, *Jaeyhung Choi, *Ishan Mor,*Goitseone Montiya, *Steven Talevski, *Yasas Thalagala	Innovative approach for NEA Exploration
Brett Biddington	Dr Rosalind Dubs and Space Industry innovation Council	The Space Industry Innovation Council: Opportunity and Challenge
Brett Biddington	Carol Oliver, Matthew Connell, Salah Sukkarieh, Michael O'Brien, and Brett Biddington	Pathways to Space: Empowering the Internet Generation
Mark Bishop	Mark A. Bishop and David Tran	Linear Dune Spacing and the Influence of Topography, Simpson Desert: An Analogue for the Understanding of Dune Formation on Earth and Titan
Mark Blair	Mark Blair	ASRI Programs in Space Education
Russell Boyce	Russell Boyce, Sean O'Byrne, Con Doolan, David Buttsworth, Allan Paull, Philip Teakle, Jerome Vethecan, Don Fry, Klaus Hannemann, Tetsuji Sunami, Gennaro Russo, Graham Candler, University of Minnesota (UMN), Stuart Kearney	The ASRP-funded project "Scramjet-based Access-to-Space Systems"
Laurie Brown	Laurie Brown	Numerical Considerations for the Simulation of Supersonic Combustion Ramjet Intakes
Daniel Bunker	D. Bunker and Dr. S. Sukkarieh	An Electrodynamic Tether System for Picosatellites
David Buttsworth	David Buttsworth and Michael Jokic	Beagle 2 Mars Lander Flight Dynamics Investigation via Experiment and Simulation
Manuel Cervera	Manuel Cervera, Trevor Harris	Modelling of the effects of ionospheric disturbances using 3D magneto-ionic raytracing techniques
Christine Charles	Christine Charles	Australian research on plasma thrusters for spacecraft propulsion
Rattanasuda Cholathat	Rattanasuda Cholathat, Linlin Ge, Xiaojing Li, Chris Rizos	Monitoring Geologic Sequestration with Radar Remote Sensing
Aditya Chopra	Aditya Chopra and Charles Lineweaver	Stars to Planets to Life: An Elemental Scheme

Name of Presenter	Names of all authors	Title
Jen Jen Chung	Jen Jen Chung and Salah Sukkarieh	High Level Risk Analysis and Decision Making for an Autonomous Mars Glider
David Cooper	David Cooper	Spaceward Bound Australia – Pilbara 2011 – and it's predecessors
Bernard Davison	Richard Samuel and Bernard Davison	ASRI Small Sounding Rocket Program Activities and Developments
Alina Donea	Alina Donea	Australian Research on Solar Physics
Thomas Fahey	Thomas Fahey	Ausroc Nano Intertank Fairing Structural Analysis and Design
Thomas Ferguson	Thomas Ferguson, Thomas Jazra, Michael Kevin Smart	Rocket Booster Design for Scramjet-Assisted Access-to-Space Vehicles
Yue Gao	Yue Gao, Craig Smith, Ben Greene	Laser Tracking of Space Debris
Jed Guinto	Jed Guinto and Lachlan Thompson	Artificial Muscles for EVA Space Glove Design
Li Guo	Li Guo, Linlin Ge, and Xiaojing Li	Integration of remotely sensed indices for land cover changes caused by the 2009 Victorian bushfires using Landsat TM imagery
Trevor Harris	Trevor Harris and Manuel Cervera	SpICE: A Program to Study Small-Scale Disturbances in the Ionosphere
Alan Harrland	Alan Harrland	Hypersonic Inlet for a Laser Powered Propulsion System and its Interaction with an Idealized Laser Induced Detonation Wave
Jason Held	Dr. Jason Held, Mr. Nigel Hoschke, Dr. Don Price	Estimating Spacecraft Structural Health Impacts Using Hybrid State Space Models
Jason Held	Dr. Jason Held, Dr. Alex Green, Mr. Michael Lamont	The Operationally Responsive Space Project
Dillon Hunt	D.C. Hunt, R.R. Boyce and A. Paull	Computational Investigation Of An Axisymmetric Scramjet Configuration
Trevor Ireland	Trevor Ireland	HAYABUSA Returns
Michael Jokic	Michael Jokic, David Buttsworth, Albert Chong	Temperature Estimation of the Hayabusa Capsule during Re-entry via direct Near Infrared Imaging
Eriita Jones	Eriita Jones, and Charles Lineweaver	Planetary Phase Models - Searching For Life
Eriita Jones	Eriita Jones, Johnathan Clarke, Frank Mills, Bruce Doran, and Charles Lineweaver	Thermal Land Cover Map of Mars (a GIS-based Study)
Chontisa Kalyanamitra	Chanapat Bhadrakom, Chontisa Kalyanamitra, Pongpaiboon Akapaiboon, Siwat Koekiatpibul, Thanit Chanjarhatpong, Toppak Chiammunchit	An AUSROC NANO
Timothy Kaufmann	Timothy Kaufmann	The Mobile Space Academy (MSA) Concept: Practical Hands-on Science & Engineering Generates Enthusiasm for Space
Lucyna Kedziora-Chudczer	Lucyna Kedziora-Chudczer	VSTAR modelling of the Jupiter clouds
Razmi Khan	Razmi Khan, Troy Eichmann, Ben Upcroft, David Buttsworth	Automated Visual Tracking and Spectral Analysis of the Hayabusa Re-entry

Name of Presenter	Names of all authors	Title
Myrtille Laas-Bourez	Myrtille Laas-Bourez, John A. Kennewell and David M. Coward	Tracking Australian Quasi-Geosynchronous Space Debris with the Zadko Telescope
Nicholas Lawrance	Nicholas R.J. Lawrance, Salah Sukkarieh	Autonomous soaring for atmospheric exploration
Andrew Layden	Andrew Layden, Iver Cairns, Peter Robinson	Changes in mode properties versus mode conversion for waves in Earth's auroral ionosphere
Brett Layden	Brett Layden, John Percival, Iver Cairns, Peter Robinson	Thermal Correction to the Rate of Second Harmonic Plasma Emission
Joseph Leach	J. H. J. Leach and Q. J. K. Tan	A late flood event in the formation of the Warrego Valles system on Mars
Charles Lineweaver	Charles Lineweaver, Aditya Chopra	The Earliest Divergences in the Tree of Life and Constraints on the Earliest Terrestrial Environments
Yonghua Liu	Y. H. Liu, B. J. Fraser, F. W. Menk	Pc2 waves observed by Cluster and Antarctic ground stations in dayside outer magnetosphere
Austyn Luke	Austyn Luke, Jackson Richards	HIGHGlide Project - Rocket Deployment of UAVs
Kenneth Lynn	Kenneth Lynn	VLF Gravity Wave Observations
Franklin Mills	F. P. Mills, B. J. Sandor, R. T. Clancy, Y. L. Yung, and M. Allen	Analytic modeling of SOx in Venus' mesosphere
Chandrakant More	Chandrakant More, Ashokkumar Sharma and Rajaram Bhonsle	Field strength measurement of VLF radio wave propagation at 19.8 KHz between Australia and India
Oscar Moze	Oscar Moze	Space Science in Italy: links between science and industry and the economic potential of the collaboration
Dave Neudegg	Dave Neudegg	IPS Radio and Space Services - The Bureau of Meteorology space weather capability.
Sean O'Byrne	Sean O'Byrne, Varun Prakash	Embedded Data Acquisition System for Hypersonic Heat Flux Measurements
Hideaki Ogawa	Hideaki Ogawa and Russell Boyce	Physical Insight into Nozzle Flow Behaviour of Axisymmetric Scramjets for Access-to-Space via Design Optimisation
John Olsen	Rhys Wolfendon, John Olsen, John Page	Simulations of the effect of electrode geometry on the performance of an electric ion rocket
Marco Ostini	Macro Ostini	Lunar Numbat - Open Source Space technologies
David Petty	D. Petty, M. Smart, V. Wheatley and S. Razzaqi	Numerical Simulations of Hypervelocity Scramjet Combustion Flows
Robert Pidgeon	Robert Pidgeon	Planetary Science in Australia
Robert Pidgeon	Robert.Pidgeon, Marion Grange and Alexander Nemchin	The Early Heavy Bombardment of the Moon at 4.34 Ga
Drew Ravalico	Drew Ravalico	The Challenges in the Design and Build of the University of Adelaide CubeSat
Sarah Razzaqi	Sarah Razzaqi, Thomas Jazra, Michael Kevin Smart	Optimisation of Scramjet-Assisted Access-to-Space Vehicles Using Oxygen Enrichment
Anthony Rea	Anthony Rea, Susan Barrell	Remote Sensing in the Bureau of Meteorology's Composite Observing System
Obaid ur Rehman	Obaid ur Rehman, Ian Petersen	Nonlinear Robust Control design for Hypersonic Flight Vehicles

Name of Presenter	Names of all authors	Title
Sam Reisenfeld	Sam Reisenfeld	Satellite Modems and Communications Technology
Chris Rizos	Andrew Dempster, Chris Rizos	SAR Formation Flying – a Second Round ASRP Project
Chris Rizos	Chris Rizos	Australia & Future Satellite Navigation Systems – the "Lucky Country"
Elyse Schinella	Elyse Schinella and Craig O'Neill	Constraints on crustal rheology from the evolution of Venus' geological landforms
Leila Norouzi Sedeh	Leila Norouzi Sedeh, Colin Waters	Properties of ULF waves observed by TIGER SuperDARN radars
Cameron Sinclair	Cameron Sinclair, Kylie Bedwell, Daniel Bettcher, Laura Brooks, Matthew Tetlow	Sensitivity and Single Failure Mode Analyses of a Six Degree of Freedom Model for a Multi-Stage Launch Vehicle
Elizabeth Smith	Elizabeth Smith, Richard Marshall, Colin Waters, Murray Sciffer	A Risk Assessment of Space Weather Effects on the Australian Power Network
Ted Steinberg	Ted Steinberg, Martin Castillo, Owen Plagens, David Lynn, Matthew Hales, Antoine Diana and Wayde Martens	Queensland University of Technology Reduced Gravity Drop Tower Facility and Research
Steven Talevski	Steven Talevski and Lachlan Thompson	Completion of the AUSROC Nano gimballing system
Ken Tan	Q. J. K. Tan, J. H. J. Leach	The Channels of Mars and the Basalt Plains of Victoria: An Investigation of Channel Formation Chronology
Mike Terkildsen	Mike Terkildsen	Australian Research on Space Weather
Sandy Tirtey	Sandy Tirtey	The SCRAMSPACE I Hypersonic Flight Experiment Feasibility Study
Paul Tregoning	Paul Tregoning	Science Results from Australian Geodesy
Paul Tregoning	Paul Tregoning, Daniel Shaddock, Craig Smith, Bob Oreb, Bruce Warrington	The GRACE Follow On mission
Franziska Ullrich	Franziska Ullrich and Salah Sukkarieh	Design Optimization of a Mars Rover Rocker-Bogie Mechanism using Genetic Algorithms
Robert Vincent	Robert Vincent	Australian Research in the Middle Atmosphere and Lower Ionosphere
Xin Wang	Xin Wang, Linlin Ge, Xiaojing Li and Michael Chang	Detection of Pastures in Southern Western Australia based on ENVISAT ASAR Dual Polarimetric Data
Xin Wang	Hai Tung Chu, Linlin Ge and Xin Wang	Using dual polarized L band SAR and optical satellite imagery for land cover classification in Southern Vietnam: Comparison and Combination
Andrew Wheeler	Andrew Wheeler, Mark A. Bishop	Quantifying Crescentic Dunefield Development Utilizing Spatial Statistics, North Polar Plains, Mars
Andrew Winstanley	Andrew Winstanley	Hybrid Rocket Autonomous Fly Home Recovery System
Stefan Würgler	Stefan Würgler and Salah Sukkarieh	Path Planning for a Planetary Rover
Hongang Yang	Hongang Yang, Brian Fraser and Murray Sciffer	Wave induced particle precipitation using test-particle simulations

Name of Presenter	Names of all authors	Title
Hannah Young	Hannah H. Young	From the backyard to beyond: investigating the cultural significance of amateur satellites
Kefei Zhang	Kefei Zhang	The Australian Space Research Program: Platform Technologies for Space, Atmosphere and Climate project
Yuanyuan Zhang	Yuanyuan Zhang, Linlin Ge and Xiaojing Li	Comparison of HPF fusion and Wavelet fusion approaches on Image Fusion of Hyperspectral image and Multispectral image
Yingxin Zuo	Yingxin Zuo and Linlin Ge	Evaluation of ALOS PALSAR Applicability to Generate Bushfires Scars Map

Posters

Posters are listed in alphabetical order of first author

Name of First Author	Names of all authors	Title
Zahra Bouya	Zahra Bouya	Sources and Sinks : A study of polar cap Pc1-2
Alan Brockman	Alan H Brockman, Myrtille Laas-Bourez, and John A Kennewell	Monitoring Space Debris in Australia
Nerida Hector	Mark A. Bishop, Nerida Hector, Andrew Wheeler	Comparative Pattern Evolution and Self-Organization for Dunefields of the Ar Rub' al Khali Sand Sea, Earth, and the North Polar Plains, Mars
Dean Hillan	Dean Hilllan, Iver Cairns, Peter Robinson	Type II Solar Radio Bursts : Extraction of shock parameters and detailed comparison of theory with observations
Timothy Kauffman	Timothy Kauffman	NASA ISS Microgravity Research Platform Open for Business
Bo Li	Bo Li Bo Li, Iver H Cairns, and Peter A Robinson	Probing Spatial Temperature Variations in Solar Corona via Type III Solar Radio Bursts
Joseph Leach	Joseph Leach, Richard Arndt, Botao Shen, Zhong Xiao Goh, Damian Yeung, Noriko Hirashima, and Zhixu Li	Site selection for long duration lunar bases
Vasili Lobzin	Vasili Lobzin,Iver Cairns, Alexander Warmuth, Roman Gorgutsa, Peter Robinson, Gottfried Mann, Valery Fomichev	Evidence for Gently Sloping Plasma Density Profiles in the Deep Corona
Seann McKibbin	Seann McKibbin	Mn-Cr systematics in meteoritic olivine measured by SHRIMP (Sensitive High-mass Resolution Ion Micro-Probe)
Franklin Mills	Manuraj Shunmuga Sundaram, Franklin Mills, Mark Allen, Yuk Yung	A modelling assessment of nitrogen oxide chemistry in the Venus mesosphere
Chandrakant More	Chandrakant More and Ashokkumar Sharma	Studies On Effect Of Solar Activities On Transequatorial Vlf Radio Wave Propagation At 19.8 Khz
Fiona Schleyer	Fiona Schleyer, Iver Cairns, Eun-Hwa Kim, Peter Robinson	Linear mode conversion of upper hybrid waves to radiation: Averaged energy conversion efficiencies, polarization, and applications to Earth's magnetosphere
Graham Steward	Graham Steward, Vasili Lobzin, Phil Wilkinson, and Dave Neudegg	Automatic Recognition of Complex Magnetic Regions on the Sun using GONG Magnetogram Images and Their Usefulness in Predicting Flares
Dion Tiu	Dion Tiu and Iver Cairns	Evidence for Reformation of the Uranian Bow Shock
Xiaofeng Wu	Xiaofeng Wu and Xueliang Bai	A Virtual Satellite Design for the University Student Satellite Program
Hongang Yang	Hongang Yang A. Bhattacharjee, Yi-Min Huang, Lei Ni, B. Rogers	Fast reconnection in high-Lundquist-number plasmas due to the shear flow and plasmoid Instability